This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number: **0 499 681 A1**

(12) **EUROPEAN PATENT APPLICATION**

(21) Application number: **91106110.9**

(22) Date of filing: **17.04.91**

(51) Int. Cl.[5]: **C12N 15/30**, A61K 39/118, G01N 33/569, C07K 15/00

Priority : IT MI91A000314.

(30) Priority: **07.02.91 IT 31491**

(43) Date of publication of application:
26.08.92 Bulletin 92/35

(84) Designated Contracting States:
AT BE CH DE DK ES FR GB GR IT LI LU NL SE

(71) Applicant: **SCLAVO S.p.A.**
Via Fiorentina 1
I-53100 Siena(IT)

(72) Inventor: **Ratti, Giulio**
Strada di Vico Alto, 6
I-53100 Siena(IT)
Inventor: **Comanducci, Maurizio**
Via Montanini, 54
I-53100 Siena(IT)
Inventor: **Tecce, Mario F.**
Via Mino Celsi, 12
I-53010 Costalfino (Siena)(IT)
Inventor: **Giuliani, Marzia M.**
Via Garibaldi, 58
I-53034 Colle val d'Elsa (Siena)(IT)

(74) Representative: **Moretti, Giorgio et al**
Notarbartolo & Gervasi s.r.l. Viale Bianca Maria, 33
I-20122 Milano(IT)

(54) **PCTD plasmid isolated form chlamydia trachomatis serotype D, its genes and proteins encoded by them; recombinant plasmids for the expression of said genes in heterologous systems as fused recombinant proteins, preparation of said recombinant proteins and their use in the formulation of vaccins and/or diagnostics.**

(57) A plasmid isolated from Clamydia trachomatis is described, which comprises 8 genes encoding proteins useful in the formulation of vaccines or diagnostic test for determining the bacterium or specific antibodies generated during C. trachomatis infections; in particular the recombinant fusion MS2-pgp3D protein is described comprising polypeptidic sequences encoded by pCT and immunogenic in the course of infections in man. A method for preparing said protein in E.coli further described.

EP 0 499 681 A1

Invention Field

This invention refers to the pCTD plasmid isolated from Chlamydia trachomatis serotype D, cloned and sequenced and to the genes present in said plasmid, to the proteins expressed by said genes, to the expression vectors containing said genes and to the microrganisms transformed by said vectors. The invention further refers to the process for the preparation of genes and of said vectors and to the use of said proteins as antigens for the preparation of polyclonal and monoclonal antibodies apt to recognize Chlamydia trachomatis and hence useful for the preparation of vaccines capable of imparting a protective immunity against infections caused by Chlamydia trachomatis and pathologic conditions deriving from said infections and for the development of diagnostic methods for the search of specific antibodies produced following C.trachomatis infections.

Prior art

Chlamydias are gram-negative bacteria, obligate intracellular parasites of eukariotic cells. Chlamydias show an extracellular infective and metabolically practically inert form, called elemental body (EB), and intracellular replicative forms called reticular bodies (RB).

The reticular bodies, after multiplication by binary fission, are transformed into elemental bodies which come out of the host cell and infect new cells.

The masses or mini-colonies of reticular and elemtal bodies inside an infected cell constitute the characteristic "inclusions" visible at the optical microscope.

Chlamydia trachomatis (C.trachomatis or CT), a bacterial species pathogenic to man, is the etiological agent of venereal lymphogranuloma (VLG), of various inflammatory patologies of the genital male and female apparatus and of trachoma, a chronic disease which affects 500 million people and can lead to blindness.

In the technical literature ca. 15 CT serotypes pathogenic to man were described and divided in two groups which differ both as to virulence and tissular tropism.

Twelve serotypes of the trachoma group (biovar) are identified as A to K and infect, in general, epithelial tissues, such as the ocular (trachoma) and uro-genital (cervicitis and urethritis) mucous membranes, and show a low virulence.

The venereal lymphogranuloma (VLG) serotypes (L_1, L_2 and L_3) cause instead an infection of the reticulo-endothelial tissue, mainly of the inguinal and femoral lymphonodi, and are highly invasive.

Urethritis and cervicitis induced by CT (A to K serotypes) when not precociously diagnosed and treated by adequate therapy, may led to a variety of chronic inflammations, such as, e.g., vaginitis, salpingities and pelvic inflammation which may resolve in sterility and extrauterine pregnancy.

Furthermore the new born from infected mothers may contract pulmonary and/or ocular infections during delivery.

For said reason it is necessary to possess adequate diagnostic methods for determining CT and formulating effective vaccines against said bacterium.

As known, factors which determine the bacterial virulence are often encoded by genes present on plasmids.

In the literature, the presence is reported, in all 15 serotypes and in the clinical isolates examined up to now, of a plasmid of ca. 7.5 Kb referred to in the present invention as pCT followed by the denomination of the bacterial serotype concerned. For example: pCTD for the plasmid isolated from serotype D, etc.

Up to now, however, no specific function or products encoded by it were associated with said plasmid.

Detailed description of the invention

A variant of the plasmid, corresponding to serotype D, was now isolated, indicated in what follows a pCTD, which comprises at least eight genes encoding for new proteins.

Figure 1a shows the nucleotidic sequence of said plasmid and 7 of the 8 protein structures expressed by said sequence. The eighth protein structure, encoded on the DNA chain complemental to the one of Fig. 1a, is shown in Fig. 1b.

Object of the present invention are thus: the cloned and sequenced pCTD plasmid, the nucleotide sequences encoding for the above named proteins, the expression vectors containing one of said sequences or fragments thereof.

Further object of the present invention are the pCTD proteins or fragments of them having immunogenic properties.

Still another object of the present invention are the fusion polypeptides comprising one of said proteins or its fragments suitable as antigens.

The present invention further refers to the preparation of said proteins and of their fragments possessing immunogenic activity or of fused polypeptides comprising said proteins.

Said proteins, their fragments or fusion polypeptides comprising said proteins or their fragments, according to the invention may be employed to determine the CT produced infections in biological samples.

Said proteins, their fragments or fusion polypeptides comprising the protein or its fragments may further be employed, according to the invention, as antigens useful in the formulation of vaccines against infections due to CT.

According to the invention, said proteins, their fragments or fusion polypeptides may be used furthermore as antigens for the preparation of poly- or mono-clonal antibodies to be used in diagnostics. In particular, the present invention relates to the pgp 3D protein encoded by the gene of the pCTD plasmid identified as ORF3D having the nucleotide sequence reported in Fig. 2, and characterized by a molecular weight of 27,802 and by the aminoacid sequence reported in Fig. 2.

According to the present invention, plasmid pCTD is obtained from the C.trachomatis GO/86 strain isolated from the urethra of a patient with non-gonococcic urethritis, and successively identified as serotype D by the immunofluorescence method described by Wang, S.P. and Grayston, J.T. [Am. J. Ophtalmol. 70; 367-374 (1970)]. The ORF3D gene may be isolated from the pCTD plasmid employing one of the known methods such as, e.g., the in vitro amplification method [Saiki, A.K. et al. Science, 239 :487-491 (1988)] using as primers synthetic oligonucleotides having a primary structure suitably derived from the sequence data shown in Figs. 1a and 1b. The thus emplified gene is then cloned in a vector placing it under the control of sequences regulating its expression.

One can similarly proceed for the other seven genes the nucleotide sequences of which are reported in Figs. 1a and 1b.

The proteins encoded by said genes are represented by the aminoacid sequences also reported in Figs. 1a and 1b.

Vectors suitable for the ends of the present invention may be plasmids with expression in host cells selected among the ones known and available commercially or at authorized collection centers.

The cells transformed by said vectors are then cultivated in a suitable culture medium in the presence of carbon-, nitrogen- and mineral salts sources, possibly in induction conditions, at a temperature and time period selected in order to obtain the production of the desired protein.

Said protein, obtainable also as fused polypeptide, constituted by a polypeptide produced by the vector fused with the protein itself, is then separated and purified from the culture medium or from the cell lysate.

According to one embodiment of the present invention, the ORF3D gene is cloned in the plasmidic E.coli pEX34a vector, a derivative of pEX29 and pEX31 described by Strebel et al. [J.Virol., 57:983-991 (1986)], following the description by Nicosia et al. in Infect. Imm. 1987, Vol.55, 963-967.

The results show the presence in the bacterial extracts of a polypeptide, indicated as MS2-pgp3D, the sequence of which is shown in Fig. 3, with a mol. weight of ca. 39 Kd, consisting i.e. of a RNA-polymerase fragment of bacteriofage MS2, produced by the expression system of ca. 11 Kd and by the protein encoded by the ORF3D gene of ca. 28 Kd.

Said polypeptide employed as antigen in a Western-Blot assay, or in immunologic assays, is recognized by antibodies present in the serum of patients with CT infection and may further be employed for the production, in laboratory animals, of mono- and poly-clonal antibodies which recognize the - and react with the corresponding pgp3 protein, in all its variants, of C.trachomatis.

In accordance with the present invention the pCTD and p03/60/MCI plasmids were deposited as ATCC N° 68314 and ATCC N° 68315 respectively.

The experimental examples that follow are illustrative and non limitative of the invention.

EXAMPLE 1

Isolation of the pCTD plasmid from C.trachomatis GO/86

C.trachomatis cells were isolated following known techniques from the urethra of a patient with non-genococcic urethritis. The strain, identified as serotype D by the micro-immunofluorescence technique described by Wang, S.P. and Grayston, J.T. [(1970), Am. J. Ophtalmol., 70: 367-374] is designated as GO/86.

The elemental bodies of said strain are then purified as described by Cevenini R. et al. [(1988), FEMS Microbiol. Letters, 56:41-46] on renografin^{R} density discontinuous gradients (E.R. Squibb & Sons, Princeton,

N.J.) according to what reported by Caldwell H.D. et al. [(1988) Infect. Immun. 31:1161-1176].

After purification, the elemental bodies (ca. 1.5 mg proteins) are lysated by incubation in 10 mM Tris-HCl, pH 8.0, 150 mM NaCl, 2mM EDTA, 0.6% SDS and 100 mg/ml K Proteinase (Boehringer) at 37°C for 3 hrs. The total nucleic acids are then extracted with phenol/chloroform, precipitated with ethanol, treated with pancreatic RNAse (250 ng/µl final concentration), further precipitated with ethanol and re-suspended in 800 µl water (365 ng/µl of DNA).

A 10 µl aliquot of said solution is then treated with 30 units (U) of BamHI restriction enzyme (Boehringer) at 37°C for 2 hrs in 20 µl (final volume) of a digestion mixture suggested by the supplier. 3 µl of the resulting digestion mixture are ligated to 100 ng plasmidic pUC8 DNA previously digested with BamHI and dephosphorilated with calf gut phosphatase. The ligase reaction is effected overnight in 20 µl buffer containing 9 U T4 DNA ligase (Boehringer) at 18°C.

The ligation mixture is then employed to transform HB101 E.coli cells made competent by a treatment with $CaCl_2$ as described by Mandel and Higa [(1970) J. Mol. Biol: 53, 54]. The transformants are selected on LB agar Medium (DIFCO) with addition of 100 µg/ml ampicillin, at 37°C overnight.

The positive clones (ampicillin resistant) (Amp^R) containing, that is, the recombinant pUC8 plasmid are transferred onto Hybond-N membranes (Amersham) and sorted by hybridization with three ^{32}P marked oligonucleotides having the following nucleotidic sequences:

1) 5'ATGGGTAAAGGGATTTTATC3'

2) 5'CTATATTAGAGCCATCTTC3'

3) 5'TCAAAGCGCTTGCACGAAG3'

The above reported oligonucleotides are synthesized by means of an automatic synthesizer (Applied Biosystem Inc. Mod. 380A) following the methods and employing the reagents recommended by the manufacturers.

Four of the six plasmids isolated from the clones found positive at the hybridization, analyzed by electrophoresis on agarose 1% gel before and after digestion with BamHI are found to consist of the pUC8 plasmid nucleotidic sequence and of a nucleotidic insert of ca. 7.5 kilobases corresponding to the isolated C.trachomatis GO/86 plasmid.

The nucleotidic sequences of said insert is determined according to the method of Sanger F. [(1977) PNAS USA 74:5463-5467] utilizing a series of suitable primers. The sequencing reactions are performed on double helix DNA employing the Sequenase Kit (U.S. Biochemical Co. Cleveland, Ohio) as recommended by the firm.

The nucleotidic sequences of the ca. 7.5 kilobases plasmid named pCTD are reported in Figs. 1a and 1b. The recombinant plasmid containing said insert is indicated as pUC8-pCTD.

EXAMPLE 2

Cloning of the DNA ORF3D segment of plasmid pCTD1D

The DNA fragment denoted as ORF3D(Fig. 2) of 792 bp is obtained through in vitro amplification according to the technique known as Polymerase Chain Reaction (PCR) described by Saiki A.K. et al. [-(1988) Science 239:487-491].

The amplification is effected utilizing ca. 10 ng of the pUC8-pCTD plasmid and employing as primers two synthetic oligonucleotides (ORF31) and (ORF3dx) having respectively the following nucleotide sequences:

- 5'CAGGGATCCATGGGAAATTCTGGTTTTT3'

BamHI

- 5'CCCCTGCAGTTAAGCGTTTGTTTGAGGT3'

Pst I

Said oligonucleotides are complemental to ORF3 regions with the addition to the respective 5' terminals of a nucleotide sequence comprising the action site of a restriction enzyme selected among the ones present in the pEX34A vector (Strebel K. et al. [(1986) J. Virol.57: 983-991] utilized for the successive cloning. In particular, the site selected for ORF31 is the one for the BamHI enzyme, while for ORF3dx is the one of the PstI enzyme.

The amplification reaction is performed employing the reagents contained in the "Geneamp" Kit (Perkin Elmer-Cetus). 25 amplification cycles are effected. Each amplification cycle consists in heating the reaction mixture to 94 °C for one minute, to 50 °C for one minute and finally to 72 °C for one minute.

At the end of the amplification reaction the mixture is extracted, in succession, with an equal volume of phenol and of a chloroform-isoamyl alcohol mixture (24:1 v/v) and then submitted to forced dialysis by means of Centricon^{R} cartridges following the producer's (Amicon) instructions.

The DNA is then precipitated by adding to the obtained solution sodium acetate 3 M, pH 5.5 (1/10 of the volume) and cold (-20 °C) ethanol (3 vols.). The DNA precipitate is dissolved in 44 µl water. To the solution, 5 µl H buffer (Boehringer) and 1 µl PSTI restriction enzyme (20 units/µl) are added and the DNA is digested at 37 °C for 2 hours.

The digestion mixture is then extracted with phenol, chloroform/isoamyl alcohol and then the DNA is precipitated with ethanol (-20 °C). The precipitate, separated by centrifugation, is suspended again in 44 µl water and then digested with 20 U BamHI in 5 µl of B buffer (Boehringer) at 37 °C for 2 hours. The digestion mixture is extracted with phenol, chloroform/isoamyl alcohol and dialyzed by Centricon^{R} cartridge.

At the same time, 10 µg of the pEX34A plasmidic vector are digested with the PstI and BamHI restriction enzymes as reported supra. The vector is dephosphorylated with alkaline phosphatase, extracted with phenol and chloroform/isoamyl alcohol, precipitated with ethanol (-20 °C) and re-suspended in 50 µl water.

1 µl (100 ng) of the vector and 2 µl (200 ng) of the amplified ORF3D segment are then ligated in 2 µl ligase buffer to which 2 µl ATP r, 1µl T4 DNA ligase (9 units/µl) are added, adding water to a total volume of 20 µl. The ligase reaction is performed at 15 °C overnight. The ligase mixture is employed to transform 200 µl of a suspension of E.coli competent cells (K12-ΔH1-Δ trp) [Remaut E. et al. (1983), Gene 22:103-113]. After treatment at 30 °C for 5 minutes, to the cell suspension 800 µl LB medium are added, followed by incubation at 30 °C for 1 hour. Aliquots of the cell suspension (10 µl, 100 µl and 690 µl) are separately plated on plates of agarized (20 g/l) LB medium containing 100 µg/mg ampicillin and kept at 30 °C overnight.

The obtained clones (Amp^{R}) are transferred to a nitrocellulose membrane on a LB agar plate with added ampicillin, grown at 30 °C overnight, and then tested for hydridization with three oligonucleotidic probes (UB35, UB36, UB18) terminally marked with ^{32}P having the following sequences:

I) 5'-ATGGGTAAAGGGATTTTATC3'

II) 5'-CTATATTAGAGCCATCTTC3'

III) 5'-TCAAAGCGCTTGCACGAAG3'

The hybridization test is performed according to known tecnique. From the colonies positive to hybridization the plasmids contained in them are prepared by minipreparation as described by Maniatis et al. (1982) and the ORF3D insert nucleotide sequence is controlled by known technique.

EXAMPLE 3

Expression of the MS2-gpg3 recombination protein

E.coli cells containing the pEX34 vector with the ORF3D insert are inoculated in duplicate in 10 ml LB medium with added 30 μg/ml ampicillin and cultivated at 30°C overnight. The procedure described by Nicosia et al. [Inf. Imm. (1987) 55:963-967] is then followed, with the provision that one of two duplicates undergoes induction of the cloned gene by treatment at 42°C, while the other does not. Two protein extracts are thus obtained, produced by the bacterium, in 7M urea buffered at pH 8, one of which corresponds to the induced cells, and the other, as a control, to the non-induced cells.

By analysis of the protein contents of both extracts by electrophoresis in SDS-polyacrylamide 15% gel according to known techniques, it is possible to deduct the presence of a protein species of 39,000 apparent mol.wt. which is present in a considerably greater amount in the induced extracts.

In the non-induced cell lysate no evidence of such a protein, but only the product of the vector alone, is found.

Said electrophoresis patterns may be analyzed by the Western Blot technique employing a monoclonal antibody (SCLAVO) specific for the 11 kd fragment generated by the pEX34 vector. In this way it is possible to demonstrate that the 39 kd band is a fusion protein containing said fragment.

EXAMPLE 4

Purification of MS2-pgp3 from E.coli K12Δ H1Δ trp extracts

The protein extract, from induced bacterial cells, re-suspended in 7M urea, is dialyzed for 15 hrs. at 4°C against a PBS buffer consisting of 0.4% KCl, 0.4% KH_2PO_4, 16% NaCl, 2.5% NaH_2PO_4.

During the dialysis a protein precipitate is obtained, which is separated by centrifuging and discarded. The sumatant is submitted to further purification by electrophoresis on preparative 12.5% acrylamide gels, and the protein band of 39,000 mol.wt. (MS2-pgp3D) is then extracted by electroelution from the gel.

The thus obtained MS2-pgp3 is precipitated by adding to the electroeluted solution 9 volumes of absolute acetone (-20°C). The protein precipitate is separated by centrifuging, re-suspended in 90% acetone, centrifuged as above, precipitated in 96% acetone and centrifuged again. The precipitate is brought to dryness in a nitrogen stream and re-suspended in 200 μl sterile PBS at a final concentration of approximately 1.5 μg/μl.

The advantage of the effected dialysis is the elimination, with this procedure, of some E.coli proteins, in particular some with a molecular weight equal or very near to the one of the desired recombinant product, which may present a considerable hinderance in the electrophoretic and/or chromatographic purification.

EXAMPLE 5

Production of polyclonal anti-MS2-pGPG3 antibodies

Utilizing the MS2-pgp3 protein, purified as in Example 4, 3 Balb/C 7-8 week old mice are immunized intraperitoneally. The immunization procedure comprises a first injection of 0.2 ml/mouse of an emulsion consisting of one part by vol. of the purified protein solution (1.5 μg/μml) and five parts of Freund complete adjuvant (FCA).

The thus inoculated protein amount is thus ca. 50 μg/mouse. After 1 week the mice are immunized with the said same emulsion, followed by a 800 μl Pristane injection. After 1 week from the second inoculation, the mice are intraperitoneally immunized with 0.2 ml of a solution similar to the first one. Finally, after two weeks from the third inoculation a booster immunization is effected. The thus induced antibodies are collected in the ascitic fluid formed after the above described treatment.

The anti MS2-pgp3 antibody titres show values comprised between 1:8000 and 1:10.000 evaluated by analysis with Western Blot containing the MS2-pgp3 protein.

The reactivity of said antibodies to the native antigen (pgp3) was evaluated according to the following methods:

- analysis with Western Blot containing total protein extracts of elemental purified CT bodies
- immunofluorescence on McCoy cells cultures infected with CT. The results of the above tests show that the anti MS2-pgp3 antibodies are able to reveal C.trachomatis inclusions in infected cells (see immunofluorescence test) and recognize a protein present in the bacterium protein extracts and having a mol.wt. of 28 kd, equivalent, that is, to the one of the protein encoded by ORF3D (see Western Blot test).

EXAMPLE 6

To the end of preparing monoclonal anti-MS2-pgp3 antibodies, the mice, immunized as above described, are sacrificed, the spleens extracted and utilized for the preparation of hybridomas operating according to the technique described by Davis L.G. [Basic methods in molecular biology - Elsevier Edit., New York (1986)]. The screening of the thus obtained hybridomas is performed as described for the polyclonal antibodies. In particular, a screening was performed with induced E.coli extracts (see Example 3) containing the MS2-pgp3 protein or the polypeptide encoded by the pEX34 vector alone; obviously, the clones were selected which produced antibodies reacting only with the recombinant product. With such pgp3-specific antibodies, results are obtained which are superimposable to the ones obtained with the above described polyclonal antibodies.

EXAMPLE 7

Serum samples from 20 patients with Chlamydia generated infections were collected. Said sera contained anti-Chlamydia antibodies with titres comprised between 128 and 512, as determined by immunofluorescence against single antigen (LGV2). 15 control sera not containing anti-Chlamydia antibodies were obtained from healty donors. Western Blots were prepared, as above described, containing the MS2-pgp3 protein. These were incubated with the sera under examination diluted 1:100 and successively with peroxidase marked rabbit (anti human IgG) immunoglobines. 16 of the 20 infected patients sera contained antibodies apt to react with MS2-pgp3. The 15 healthy control sera did not give any reaction with said protein.

Claims

1. pCTD plasmid isolated from Chlamydia trachomatis serotype D characterized by the following nucleotidic sequence:

```
        10                  30                  50
ATATTCATATTCTGTTGCCAGAAAAAACACCTTTAGGCTATATTAGAGCCATCTTCTTTG

        70                  90                 110
AAGCGTTGTCTTCTCGAGAAGATTTATCGTACGCAAATATCATCTTTGCGGTTGCGTGTC

       130                 150                 170
CTGTGACCTTCATTATGTCGGAGTCTGAGCACCCTAGGCGTTTGTACTCCGTCACAGCGG

       190                 210                 230
TTGCTCGAAGCACGTGCGGGGTTATTTTAAAAGGGATTGCAGCTTGTAGTCCTGCTTGAG

       250                 270                 290
AGAACGTGCGGGCGATTTGCCTTAACCCCACCATTTTTCCGGAGCGAGTTACGAAGACAA

       310                 330                 350
AACCTCTTCGTTGACCGATGTACTCTTGTAGAAAGTGCATAAACTTCTGAGGATAAGTTA

       370                 390                 410
TAATAATCCTCTTTTCTGTCTGACGGTTCTTAAGCTGGGAGAAAGAAATGGTAGCTTGTT

       430                 450                 470
GGAAACAAATCTGACTAATCTCCAAGCTTAAGACTTCAGAGGAGCGTTTACCTCCTTGGA

       490                 510                 530
GCATTGTCTGGGCGATCAACCAATCCCGGGCATTGATTTTTTTTAGCTCTTTTAGGAAGG

       550                 570                 590
ATGCTGTTTGCAAACTGTTCATCGCATCCGTTTTTACTATTTCCCTGGTTTTAAAAAATG

       610                 630                 650
TTCGACTATTTTCTTGTTTAGAAGGTTGCGCTATAGCGACTATTCCTTGAGTCATCCTGT

       670                 690                 710
TTAGGAATCTTGTTAAGGAAATATAGCTTGCTGCTCGAACTTGTTTAGTACCTTCGGTCC

       730                 750                 770
AAGAAGTCTTGGCAGAGGAAACTTTTTTAATCGCATCTAGGATTAGATTATGATTTAAAA

       790                 810                 830
GGGAAAACTCTTGCAGATTCATATCCAAGGACAATAGACCAATCTTTTCTAAAGACAAAA

       850                 870                 890
AAGATCCTCGATATGATCTACAAGTATGTTTGTTGAGTGATGCGGTCCAATGCATAATAA

       910                 930                 950
CTTCGAATAAGGAGAAGCTTTTCATGCGTTTCCAATAGGATTCTTGGCGAATTTTTAAAA

       970                 990                1010
CTTCCTGATAAGACTTTTCACTATATTCTAACGACATTTCTTGCTGCAAAGATAAAATCC

      1030                1050                1070
CTTTACCCATGAAATCCCTCGTGATATAACCTATCCGTAAAATGTCCTGATTAGTGAAAT

      1090                1110                1130
AATCAGGTTGTTAACAGGATAGCACGCTCGGTATTTTTTTATATAAACATGAAAACTCGT
                                    ORF1 >>  MetLysThrArg
```

```
      1150                1170                1190
TCCGAAATAGAAAATCGCATGCAAGATATCGAGTATGCGTTGTTAGGTAAAGCTCTGATA
SerGluIleGluAsnArgMetGlnAspIleGluTyrAlaLeuLeuGlyLysAlaLeuIle

      1210                1230                1250
TTTGAAGACTCTACTGAGTATATTCTGAGGCAGCTTGCTAATTATGAGTTTAAGTGTTCT
PheGluAspSerThrGluTyrIleLeuArgGlnLeuAlaAsnTyrGluPheLysCysSer

      1270                1290                1310
CATCATAAAAACATATTCATAGTATTTAAACACTTAAAAGACAATGGATTACCTATAACT
HisHisLysAsnIlePheIleValPheLysHisLeuLysAspAsnGlyLeuProIleThr

      1330                1350                1370
GTAGACTCGGCTTGGGAAGAGCTTTTGCGGCGTCGTATCAAAGATATGGACAAATCGTAT
ValAspSerAlaTrpGluGluLeuLeuArgArgArgIleLysAspMetAspLysSerTyr

      1390                1410                1430
CTCGGGTTAATGTTGCATGATGCTTTATCAAATGACAAGCTTAGATCCGTTTCTCATACG
LeuGlyLeuMetLeuHisAspAlaLeuSerAsnAspLysLeuArgSerValSerHisThr

      1450                1470                1490
GTTTTCCTCGATGATTTGAGCGTGTGTAGCGCTGAAGAAAATTTGAGTAATTTCATTTTC
ValPheLeuAspAspLeuSerValCysSerAlaGluGluAsnLeuSerAsnPheIlePhe

      1510                1530                1550
CGCTCGTTTAATGAGTACAATGAAAATCCATTGCGTAGATCTCCGTTTCTATTGCTTGAG
ArgSerPheAsnGluTyrAsnGluAsnProLeuArgArgSerProPheLeuLeuLeuGlu

      1570                1590                1610
CGTATAAAGGGAAGGCTTGATAGTGCTATAGCAAAGACTTTTTCTATTCGCAGCGCTAGA
ArgIleLysGlyArgLeuAspSerAlaIleAlaLysThrPheSerIleArgSerAlaArg

      1630                1650                1670
GGCCGGTCTATTTATGATATATTCTCACAGTCAGAAATTGGAGTGCTGGCTCGTATAAAA
GlyArgSerIleTyrAspIlePheSerGlnSerGluIleGlyValLeuAlaArgIleLys

      1690                1710                1730
AAAAGACGAGTAGCGTTCTCTGAGAATCAAAATTCTTTCTTTGATGGCTTCCCAACAGGA
LysArgArgValAlaPheSerGluAsnGlnAsnSerPhePheAspGlyPheProThrGly

      1750                1770                1790
TACAAGGATATTGATGATAAAGGAGTTATCTTAGCTAAAGGTAATTTCGTGATTATAGCA
TyrLysAspIleAspAspLysGlyValIleLeuAlaLysGlyAsnPheValIleIleAla

      1810                1830                1850
GCTAGACCATCTATAGGGAAAACAGCTTTAGCTATAGACATGGCGATAAATCTTGCGGTT
AlaArgProSerIleGlyLysThrAlaLeuAlaIleAspMetAlaIleAsnLeuAlaVal

      1870                1890                1910
ACTCAACAGCGTAGAGTTGGTTTCCTATCTCTAGAAATGAGCGCAGGTCAAATTGTTGAG
ThrGlnGlnArgArgValGlyPheLeuSerLeuGluMetSerAlaGlyGlnIleValGlu

      1930                1950                1970
CGGATTATTGCTAATTTAACAGGAATATCTGGTGAAAAATTACAAAGAGGGGATCTCTCT
ArgIleIleAlaAsnLeuThrGlyIleSerGlyGluLysLeuGlnArgGlyAspLeuSer
```

```
      1990                2010                2030
AAAGAAGAATTATTCCGAGTAGAAGAAGCTGGAGAAACGGTTAGAGAATCACATTTTTAT
LysGluGluLeuPheArgValGluGluAlaGlyGluThrValArgGluSerHisPheTyr

      2050                2070                2090
ATCTGCAGTGATAGTCAGTATAAGCTTAACTTAATCGCGAATCAGATCCGGTTGCTGAGA
IleCysSerAspSerGlnTyrLysLeuAsnLeuIleAlaAsnGlnIleArgLeuLeuArg

      2110                2130                2150
AAAGAAGATCGAGTAGACGTAATATTTATCGATTACTTGCAGTTGATCAACTCATCGGTT
LysGluAspArgValAspValIlePheIleAspTyrLeuGlnLeuIleAsnSerSerVal

      2170                2190                2210
GGAGAAAATCGTCAAAATGAAATAGCAGATATATCTAGAACCTTAAGAGGTTTAGCCTCA
GlyGluAsnArgGlnAsnGluIleAlaAspIleSerArgThrLeuArgGlyLeuAlaSer

      2230                2250                2270
GAGCTAAACATTCCTATAGTTTGTTTATCCCAACTATCTAGAAAAGTTGAGGATAGAGCA
GluLeuAsnIleProIleValCysLeuSerGlnLeuSerArgLysValGluAspArgAla

      2290                2310                2330
AATAAAGTTCCCATGCTTTCAGATTTGCGAGACAGCGGTCAAATAGAGCAAGACGCAGAT
AsnLysValProMetLeuSerAspLeuArgAspSerGlyGlnIleGluGlnAspAlaAsp

      2350                2370                2390
GTGATTTTGTTTATCAATAGGAAGGAATCGTCTTCTAATTGTGAGATAACTGTTGGGAAA
ValIleLeuPheIleAsnArgLysGluSerSerSerAsnCysGluIleThrValGlyLys

      2410                2430                2450
AATAGACATGGATCGGTTTTCTCTTCGGTATTACATTTCGATCCAAAAATTAGTAAATTC
AsnArgHisGlySerValPheSerSerValLeuHisPheAspProLysIleSerLysPhe

      2470                2490                2510
TCCGCTATTAAAAAAGTATGGTAAATTATAGTAACTGCCACTTCATCAAAAGTCCTATCC
SerAlaIleLysLysValTrpEnd
          ORF2 >> MetValAsnTyrSerAsnCysHisPheIleLysSerProIleH

      2530                2550                2570
ACCTTGAAAATCAGAAGTTTGGAAGAAGACCTGGTCAATCTATTAAGATATCTCCCAAAT
isLeuGluAsnGlnLysPheGlyArgArgProGlyGlnSerIleLysIleSerProLysL

      2590                2610                2630
TGGCTCAAAATGGGATGGTAGAAGTTATAGGTCTTGATTTTCTTTCATCTCATTACCATG
euAlaGlnAsnGlyMetValGluValIleGlyLeuAspPheLeuSerSerHisTyrHisA

      2650                2670                2690
CATTAGCAGCTATCCAAAGATTACTGACCGCAACGAATTACAAGGGGAACACAAAAGGGG
laLeuAlaAlaIleGlnArgLeuLeuThrAlaThrAsnTyrLysGlyAsnThrLysGlyV

      2710                2730                2750
TTGTTTTATCCAGAGAATCAAATAGTTTTCAATTTGAAGGATGGATACCAAGAATCCGTT
alValLeuSerArgGluSerAsnSerPheGlnPheGluGlyTrpIleProArgIleArgP

      2770                2790                2810
TTACAAAAACTGAATTCTTAGAGGCTTATGGAGTTAAGCGGTATAAAACATCCAGAAATA
heThrLysThrGluPheLeuGluAlaTyrGlyValLysArgTyrLysThrSerArgAsnL
```

```
      2830                2850                2870
AGTATGAGTTTAGTGGAAAAGAAGCTGAAACTGCTTTAGAAGCCTTATACCATTTAGGAC
ysTyrGluPheSerGlyLysGluAlaGluThrAlaLeuGluAlaLeuTyrHisLeuGlyH

      2890                2910                2930
ATCAACCGTTTTTAATAGTGGCAACTAGAACTCGATGGACTAATGGAACACAAATAGTAG
isGlnProPheLeuIleValAlaThrArgThrArgTrpThrAsnGlyThrGlnIleValA

      2950                2970                2990
ACCGTTACCAAACTCTTTCTCCGATCATTAGGATTTACGAAGGATGGGAAGGTTTAACTG
spArgTyrGlnThrLeuSerProIleIleArgIleTyrGluGlyTrpGluGlyLeuThrA

      3010                3030                3050
ACGAAGAAAATATAGATATAGACTTAACACCTTTTAATTCACCACCTACACGGAAACATA
spGluGluAsnIleAspIleAspLeuThrProPheAsnSerProProThrArgLysHisL

      3070                3090                3110
AAGGGTTCGTTGTAGAGCCATGTCCTATCTTGGTAGATCAAATAGAATCCTACTTTGTAA
ysGlyPheValValGluProCysProIleLeuValAspGlnIleGluSerTyrPheValI

      3130                3150                3170
TCAAGCCTGCAAATGTATACCAAGAAATAAAAATGCGTTTCCCAAATGCATCAAAGTATG
leLysProAlaAsnValTyrGlnGluIleLysMetArgPheProAsnAlaSerLysTyrA

      3190                3210                3230
CTTACACATTTATCGACTGGGTGATTACAGCAGCTGCGAAAAAGAGACGAAAATTAACTA
laTyrThrPheIleAspTrpValIleThrAlaAlaAlaLysLysArgArgLysLeuThrL

      3250                3270                3290
AGGATAATTCTTGGCCAGAAAACTTGTTATTAAACGTTAACGTTAAAAGTCTTGCATATA
ysAspAsnSerTrpProGluAsnLeuLeuLeuAsnValAsnValLysSerLeuAlaTyrI

      3310                3330                3350
TTTTAAGGATGAATCGGTACATCTGTACAAGGAACTGGAAAAAAATCGAGTTAGCTATCG
leLeuArgMetAsnArgTyrIleCysThrArgAsnTrpLysLysIleGluLeuAlaIleA

      3370                3390                3410
ATAAATGTATAGAAATCGCCATTCAGCTTGGCTGGTTATCTAGAAGAAAACGCATTGAAT
spLysCysIleGluIleAlaIleGlnLeuGlyTrpLeuSerArgArgLysArgIleGluP

      3430                3450                3470
TTCTGGATTCTTCTAAACTCTCTAAAAAAGAAATTCTATATCTAAATAAAGAGCGCTTTG
heLeuAspSerSerLysLeuSerLysLysGluIleLeuTyrLeuAsnLysGluArgPheG

      3490                3510                3530
AAGAAATAACTAAGAAATCTAAAGAACAAATGGAACAATTAGAACAAGAATCTATTAATT
luGluIleThrLysLysSerLysGluGlnMetGluGlnLeuGluGlnGluSerIleAsnE

      3550                3570                3590
AATAGCAAGCTTGAAACTAAAAACCTAATTTATTTAAAGCTCAAAATAAAAAAGAGTTTT
nd

      3610                3630                3650
AAAATGGGAAATTCTGGTTTTTATTTGTATAACACTGAAAACTGCGTCTTTGCTGATAAT
ORF3>> MetGlyAsnSerGlyPheTyrLeuTyrAsnThrGluAsnCysValPheAlaAspAsn

      3670                3690                3710
ATCAAAGTTGGGCAAATGACAGAGCCGCTCAAGGACCAGCAAATAATCCTTGGGACAACA
IleLysValGlyGlnMetThrGluProLeuLysAspGlnGlnIleIleLeuGlyThrThr
```

```
      3730                3750                3770
TCAACACCTGTCGCAGCCAAAATGACAGCTTCTGATGGAATATCTTTAACAGTCTCCAAT
SerThrProValAlaAlaLysMetThrAlaSerAspGlyIleSerLeuThrValSerAsn

      3790                3810                3830
AATTCATCAACCAATGCTTCTATTACAATTGGTTTGGATGCGGAAAAAGCTTACCAGCTT
AsnSerSerThrAsnAlaSerIleThrIleGlyLeuAspAlaGluLysAlaTyrGlnLeu

      3850                3870                3890
ATTCTAGAAAAGTTGGGAGATCAAATTCTTGATGGAATTGCTGATACTATTGTTGATAGT
IleLeuGluLysLeuGlyAspGlnIleLeuAspGlyIleAlaAspThrIleValAspSer

      3910                3930                3950
ACAGTCCAAGATATTTTAGACAAAATCAAAACAGACCCTTCTCTAGGTTTGTTGAAAGCT
ThrValGlnAspIleLeuAspLysIleLysThrAspProSerLeuGlyLeuLeuLysAla

      3970                3990                4010
TTTAACAACTTTCCAATCACTAATAAAATTCAATGCAACGGGTTATTCACTCCCAGTAAC
PheAsnAsnPheProIleThrAsnLysIleGlnCysAsnGlyLeuPheThrProSerAsn

      4030                4050                4070
ATTGAAACTTTATTAGGAGGAACTGAAATAGGAAAATTCACAGTCACACCCAAAAGCTCT
IleGluThrLeuLeuGlyGlyThrGluIleGlyLysPheThrValThrProLysSerSer

      4090                4110                4130
GGGAGCATGTTCTTAGTCTCAGCAGATATTATTGCATCAAGAATGGAAGGCGGCGTTGTT
GlySerMetPheLeuValSerAlaAspIleIleAlaSerArgMetGluGlyGlyValVal

      4150                4170                4190
CTAGCTTTGGTACGAGAAGGTGATTCTAAGCCCTGCGCGATTAGTTATGGATACTCATCA
LeuAlaLeuValArgGluGlyAspSerLysProCysAlaIleSerTyrGlyTyrSerSer

      4210                4230                4250
GGCATTCCTAATTTATGTAGTCTAAGAACCAGTATTACTAATACAGGATTGACTCCGACA
GlyIleProAsnLeuCysSerLeuArgThrSerIleThrAsnThrGlyLeuThrProThr

      4270                4290                4310
ACGTATTCATTACGTGTAGGCGGTTTAGAAAGCGGTGTGGTATGGGTTAATGCCCTTTCT
ThrTyrSerLeuArgValGlyGlyLeuGluSerGlyValValTrpValAsnAlaLeuSer

      4330                4350                4370
AATGGCAATGATATTTTAGGAATAACAAATACTTCTAATGTATCTTTTTTAGAGGTAATA
AsnGlyAsnAspIleLeuGlyIleThrAsnThrSerAsnValSerPheLeuGluValIle

      4390                4410                4430
CCTCAAACAAACGCTTAAACAATTTTTATTGGATTTTTCTTATAGGTTTTATATTTAGAG
ProGlnThrAsnAlaEnd

      4450                4470                4490
AAAACAGTTCGAATTACGGGGTTTGTTATGCAAAATAAAAGAAAAGTGAGGGACGATTTT
                    ORF4 >> MetGlnAsnLysArgLysValArgAspAspPhe

      4510                4530                4550
ATTAAAATTGTTAAAGATGTGAAAAAAGATTTCCCCGAATTAGACCTAAAAATACGAGTA
IleLysIleValLysAspValLysLysAspPheProGluLeuAspLeuLysIleArgVal

      4570                4590                4610
AACAAGGAAAAAGTAACTTTCTTAAATTCTCCCTTAGAACTCTACCATAAAAGTGTCTCA
AsnLysGluLysValThrPheLeuAsnSerProLeuGluLeuTyrHisLysSerValSer
```

```
          4630                    4650                    4670
CTAATTCTAGGACTGCTTCAACAAATAGAAAACTCTTTAGGATTATTCCCAGACTCTCCT
LeuIleLeuGlyLeuLeuGlnGlnIleGluAsnSerLeuGlyLeuPheProAspSerPro

          4690                    4710                    4730
GTTCTTGAAAAATTAGAGGATAACAGTTTAAAGCTAAAAAAGGCTTTGATTATGCTTATC
ValLeuGluLysLeuGluAspAsnSerLeuLysLeuLysLysAlaLeuIleMetLeuIle

          4750                    4770                    4790
TTGTCTAGAAAAGACATGTTTTCCAAGGCTGAATAGACAACTTACTCTAACGTTGGAGTT
LeuSerArgLysAspMetPheSerLysAlaGluEnd

          4810                    4830                    4850
GATTTGCACACCTTAGTTTTTTGCTCTTTTAAGGGAGGAACTGGAAAAACAACACTTTCT
ORF5 >> LeuHisThrLeuValPheCysSerPheLysGlyGlyThrGlyLysThrThrLeuSer
          4870                    4890                    4910
CTAAACGTGGGATGCAACTTGGCCCAATTTTTAGGGAAAAAAGTGTTACTTGCTGACCTA
LeuAsnValGlyCysAsnLeuAlaGlnPheLeuGlyLysLysValLeuLeuAlaAspLeu

          4930                    4950                    4970
GACCCGCAATCCAATTTATCTTCTGGATTGGGGGCTAGTGTCAGAAGTGACCAAAAAGGC
AspProGlnSerAsnLeuSerSerGlyLeuGlyAlaSerValArgSerAspGlnLysGly

          4990                    5010                    5030
TTGCACGACATAGTATACACATCAAACGATTTAAAATCAATCATTTGCGAAACAAAAAAA
LeuHisAspIleValTyrThrSerAsnAspLeuLysSerIleIleCysGluThrLysLys

          5050                    5070                    5090
GATAGTGTGGACCTAATTCCTGCATCATTTTCATCCGAACAGTTTAGAGAATTGGATATT
AspSerValAspLeuIleProAlaSerPheSerSerGluGlnPheArgGluLeuAspIle

          5110                    5130                    5150
CATAGAGGACCTAGTAACAACTTAAAGTTATTTCTGAATGAGTACTGCGCTCCTTTTTAT
HisArgGlyProSerAsnAsnLeuLysLeuPheLeuAsnGluTyrCysAlaProPheTyr

          5170                    5190                    5210
GACATCTGCATAATAGACACTCCACCTAGCCTAGGAGGGTTAACGAAAGAAGCTTTTGTT
AspIleCysIleIleAspThrProProSerLeuGlyGlyLeuThrLysGluAlaPheVal

          5230                    5250                    5270
GCAGGAGACAAATTAATTGCTTGTTTAACTCCAGAACCTTTTTCTATTCTAGGGTTACAA
AlaGlyAspLysLeuIleAlaCysLeuThrProGluProPheSerIleL euGly LeuGln

          5290                    5310                    5330
AAGATACGTGAATTCTTAAGTTCGGTCGGAAAACCTGAAGAAGAACACATTCTTGGAATA
LysIleArgGluPheLeuSerSerValGlyLysProGluGluGluHisIleLeuGlyIle

          5350                    5370                    5390
GCTTTGTCTTTTTGGGATGATCGTAACTCGACTAACCAAATGTATATAGACATTATCGAG
AlaLeuSerPheTrpAspAspArgAsnSerThrAsnGlnMetTyrIleAspIleIleGlu

          5410                    5430                    5450
TCTATTTACAAAAACAAGCTTTTTTCAACAAAAATTCGTCGAGATATTTCTCTCAGCCGT
SerIleTyrLysAsnLysLeuPheSerThrLysIleArgArgAspIleSerLeuSerArg

          5470                    5490                    5510
TCTCTTCTTAAAGAAGATTCTGTAGCTAATGTCTATCCAAATTCTAGGGCCGCAGAAGAT
SerLeuLeuLysGluAspSerValAlaAsnValTyrProAsnSerArgAlaAlaGluAsp
```

```
      5530                  5550                  5570
ATTCTGAAGTTAACGCATGAAATAGCAAATATTTTGCATATCGAATATGAACGAGATTAC
IleLeuLysLeuThrHisGluIleAlaAsnIleLeuHisIleGluTyrGluArgAspTyr

      5590                  5610                  5630
TCTCAGAGGACAACGTGAACAAACTAAAAAAAGAAGCGGATGTCTTTTTTAAAAAAAATC
SerGlnArgThrThrEnd
      ORF6 >> ValAsnLysLeuLysLysGluAlaAspValPhePheLysLysAsnG

      5650                  5670                  5690
AAACTGCCGCTTCTCTAGATTTTAAGAAGACGCTTCCCTCCATTGAACTATTCTCAGCAA
lnThrAlaAlaSerLeuAspPheLysLysThrLeuProSerIleGluLeuPheSerAlaT

      5710                  5730                  5750
CTTTGAATTCTGAGGAAAGTCAGAGTTTGGATCGATTATTTTTATCAGAGTCCCAAAACT
hrLeuAsnSerGluGluSerGlnSerLeuAspArgLeuPheLeuSerGluSerGlnAsnT

      5770                  5790                  5810
ATTCGGATGAAGAATTTTATCAAGAAGACATCCTAGCGGTAAAACTGCTTACTGGTCAGA
yrSerAspGluGluPheTyrGlnGluAspIleLeuAlaValLysLeuLeuThrGlyGlnI

      5830                  5850                  5870
TAAAATCCATACAGAAGCAACACGTACTTCTTTTAGGAGAAAAAATCTATAATGCTAGAA
leLysSerIleGlnLysGlnHisValLeuLeuLeuGlyGluLysIleTyrAsnAlaArgL

      5890                  5910                  5930
AAATCCTGAGTAAGGATCACTTCTCCTCAACAACTTTTTCATCTTGGATAGAGTTAGTTT
ysIleLeuSerLysAspHisPheSerSerThrThrPheSerSerTrpIleGluLeuValP

      5950                  5970                  5990
TTAGAACTAAGTCTTCTGCTTACAATGCTCTTGCATATTACGAGCTTTTTATAAACCTCC
heArgThrLysSerSerAlaTyrAsnAlaLeuAlaTyrTyrGluLeuPheIleAsnLeuP

      6010                  6030                  6050
CCAACCAAACTCTACAAAAAGAGTTTCAATCGATCCCCTATAAATCCGCATATATTTTGG
roAsnGlnThrLeuGlnLysGluPheGlnSerIleProTyrLysSerAlaTyrIleLeuA

      6070                  6090                  6110
CCGCTAGAAAAGGCGATTTAAAAACCAAGGTCGATGTGATAGGGAAAGTATGTGGAATGT
laAlaArgLysGlyAspLeuLysThrLysValAspValIleGlyLysValCysGlyMetS

      6130                  6150                  6170
CGAACTCATCGGCGATAAGGGTGTTGGATCAATTTCTTCCTTCATCTAGAAACAAAGACG
erAsnSerSerAlaIleArgValLeuAspGlnPheLeuProSerSerArgAsnLysAspV

      6190                  6210                  6230
TTAGAGAAACGATAGATAAGTCTGATTCAGAGAAGAATCGCCAATTATCTGATTTCTTAA
alArgGluThrIleAspLysSerAspSerGluLysAsnArgGlnLeuSerAspPheLeuI

      6250                  6270                  6290
TAGAGATACTTCGCATCATGTGTTCCGGAGTTTCTTTGTCCTCCTATAACGAAAATCTTC
leGluIleLeuArgIleMetCysSerGlyValSerLeuSerSerTyrAsnGluAsnLeuL

      6310                  6330                  6350
TACAACAGCTTTTTGAACTTTTTAAGCAAAAGAGCTGATCCTCCGTCAGCTCATATATAT
euGlnGlnLeuPheGluLeuPheLysGlnLysSerEnd
```

```
      6370                6390                6410
ATATCTATTATATATATATATTTAGGGATTTGATTTCACGAGAGAGATTTGCAACTCTTG

      6430                6450                6470
GTGGTAGACTTTGCAACTCTTGGTGGTAGACTTTGCAACTCTTGGTGGTAGACTTTGCAA

      6490                6510                6530
CTCTTGGTGGTAGACTTGGTCATAATGGACTTTTGTTAAAAAATTTATTAAAATCTTAGA

      6550                6570                6590
GCTCCGATTTTGAATAGCTTTGGTTAAGAAAATGGGCTCGATGGCTTTCCATAAAAGTAG
            ORF7 >> LeuValLysLysMetGlySerMetAlaPheHisLysSerAr

      6610                6630                6650
ATTGTTTTTAACTTTTGGGGACGCGTCGGAAATTTGGTTATCTACTTTATCTTATCTAAC
gLeuPheLeuThrPheGlyAspAlaSerGluIleTrpLeuSerThrLeuSerTyrLeuTh

      6670                6690                6710
TAGAAAAAATTATGCGTCTGGGATTAACTTTCTTGTTTCTTTAGAGATTCTGGATTTATC
rArgLysAsnTyrAlaSerGlyIleAsnPheLeuValSerLeuGluIleLeuAspLeuSe

      6730                6750                6770
GGAAACCTTGATAAAGGCTATTTCTCTTGACCACAGCGAATCTTTGTTTAAAATCAAGTC
rGluThrLeuIleLysAlaIleSerLeuAspHisSerGluSerLeuPheLysIleLysSe

      6790                6810                6830
TCTAGATGTTTTTAATGGAAAAGTTGTTTCAGAGGCATCTAAACAGGCTAGAGCGGCATG
rLeuAspValPheAsnGlyLysValValSerGluAlaSerLysGlnAlaArgAlaAlaCy

      6850                6870                6890
CTACATATCTTTCACAAAGTTTTTGTATAGATTGACCAAGGGATATATTAAACCCGCTAT
sTyrIleSerPheThrLysPheLeuTyrArgLeuThrLysGlyTyrIleLysProAlaIl

      6910                6930                6950
TCCATTGAAAGATTTTGGAAACACTACATTTTTTAAAATCCGAGACAAAATCAAAACAGA
eProLeuLysAspPheGlyAsnThrThrPhePheLysIleArgAspLysIleLysThrGl

      6970                6990                7010
ATCGATTTCTAAGCAGGAATGGACAGTTTTTTTTGAAGCGCTCCGGATAGTGAATTATAG
uSerIleSerLysGlnGluTrpThrValPhePheGluAlaLeuArgIleValAsnTyrAr

      7030                7050                7070
AGACTATTTAATCGGTAAATTGATTGTACAAGGGATCCGTAAGTTAGACGAAATTTTGTC
gAspTyrLeuIleGlyLysLeuIleValGlnGlyIleArgLysLeuAspGluIleLeuSe

      7090                7110                7130
TTTGCGCACAGACGATCTATTTTTTGCATCCAATCAGATTTCCTTTCGCATTAAAAAAAG
rLeuArgThrAspAspLeuPhePheAlaSerAsnGlnIleSerPheArgIleLysLysAr

      7150                7170                7190
ACAGAATAAAGAAACCAAAATTCTAATCACATTTCCTATCAGCTTAATGGAAGAGTTGCA
gGlnAsnLysGluThrLysIleLeuIleThrPheProIleSerLeuMetGluGluLeuGl

      7210                7230                7250
AAAATACACTTGTGGGAGAAATGGGAGAGTATTTGTTTCTAAAATAGGGATTCCTGTAAC
nLysTyrThrCysGlyArgAsnGlyArgValPheValSerLysIleGlyIleProValTh
```

```
      7270                 7290                 7310
AACAAGTCAGGTTGCGCATAATTTTAGGCTTGCAGAGTTCCATAGTGCTATGAAAATAAA
rThrSerGlnValAlaHisAsnPheArgLeuAlaGluPheHisSerAlaMetLysIleLy

      7330                 7350                 7370
AATTACTCCCAGAGTACTTCGTGCAAGCGCTTTGATTCATTTAAAGCAAATAGGATTAAA
sIleThrProArgValLeuArgAlaSerAlaLeuIleHisLeuLysGlnIleGlyLeuLy

      7390                 7410                 7430
AGATGAGGAAATCATGCGTATTTCCTGTCTTTCATCGAGACAAAGTGTGTGTTCTTATTG
sAspGluGluIleMetArgIleSerCysLeuSerSerArgGlnSerValCysSerTyrCy

      7450                 7470                 7490
TTCTGGGGAAGAGGTAATTCCTCTAGTACAAACACCCACAATATTGTGATATAATTAAAA
sSerGlyGluGluValIleProLeuValGlnThrProThrIleLeuEnd

TT
```

2. pGO plasmid constituted by the pUC8 recombinant plasmid containing an insert corresponding to the nucleotidic sequence as per claim 1, cloned in the Bam H1 site.

3. Escherichia coli transformed with the plasmid according to claim 2 and deposited as ATCC 68314.

4. ORF1D gene characterized by the nucleotidic sequence comprised between 1129 and 2481 in the nucleotidic sequence according to claim 1.

5. ORF2D gene characterized by the nucleotidic sequence comprised between 2480 and 3539 in the nucleotidic sequence according to claim 1.

6. ORF3D gene characterized by the nucleotidic sequence comprised between 3604 and 4395 in the nucleotidic sequence according to claim 1.

7. ORF4D gene characterized by the nucleotidic sequence comprised between 4468 and 4773 in the nucleotidic sequence according to claim 1.

8. ORF5D gene characterized by the nucleotidic sequence comprised between 4804 and 5595 in the nucleotidic sequence according to claim 1.

9. ORF6D gene characterized by the nucleotidic sequence comprised between 5595 and 6335 in the nucleotidic sequence according to claim 1.

10. ORF7D gene characterized by the nucleotidic sequence comprised between 6560 and 7486 in the nucleotidic sequence according to claim 1.

11. ORF8D gene characterized by the nucleotidic sequence complemental to the one comprised between 41 and 1030 in the nucleotidic sequence according to claim 1.

12. Protein expressed by the gene according to claim 4 and characterized by the following aminoacid sequence:

pgp1:

MetLysThrArgSerGluIleGluAsnArgMetGlnAspIleGluTyrAlaLeuLeuGly
LysAlaLeuIlePheGluAspSerThrGluTyrIleLeuArgGlnLeuAlaAsnTyrGlu
PheLysCysSerHisHisLysAsnIlePheIleValPheLysHisLeuLysAspAsnGly
LeuProIleThrValAspSerAlaTrpGluGluLeuLeuArgArgArgIleLysAspMet
AspLysSerTyrLeuGlyLeuMetLeuHisAspAlaLeuSerAsnAspLysLeuArgSer
ValSerHisThrValPheLeuAspAspLeuSerValCysSerAlaGluGluAsnLeuSer
AsnPheIlePheArgSerPheAsnGluTyrAsnGluAsnProLeuArgArgSerProPhe
LeuLeuLeuGluGlyArgSerIleTyrAspIlePheSerGlnSerGluIleGlyValLeu
AlaArgIleLysLysArgArgValAlaPheSerGluAsnGlnAsnSerPhePheAspGly
PheProThrGlyTyrLysAspIleAspAspLysGlyValIleLeuAlaLysGlyAsnPhe
ValIleIleAlaAlaArgProSerIleGlyLysThrAlaLeuAlaIleAspMetAlaIle
AsnLeuAlaValThrGlnGlnArgArgValGlyPheLeuSerLeuGluMetSerAlaGly
GlnIleValGluArgIleIleAlaAsnLeuThrGlyIleSerGlyGluLysLeuGlnArg

GlyAspLeuSerLysGluGluLeuPheArgValGluGluAlaGlyGluThrValArgGlu
SerHisPheTyrIleCysSerAspSerGlnTyrLysLeuAsnLeuIleAlaAsnGlnIle
ArgLeuLeuArgLysGluAspArgValAspValIlePheIleAspTyrLeuGlnLeuIle
AsnSerSerValGlyGluAsnArgGlnAsnGluIleAlaAspIleSerArgThrLeuArg
GlyLeuAlaSerGluLeuAsnIleProIleValCysLeuSerGlnLeuSerArgLysVal
GluAspArgAlaAsnLysValProMetLeuSerAspLeuArgAspSerGlyGlnIleGlu
GlnAspAlaAspValIleLeuPheIleAsnArgLysGluSerSerSerAsnCysGluIle
ThrValGlyLysAsnArgHisGlySerValPheSerSerValLeuHisPheAspProLys
IleSerLysPheSerAlaIleLysLysValTrpEnd

or parts of it.

13. Protein expressed by the gene according to claim 5 and characterized by the following aminoacid sequence:

pgp2 :

MetValAsnTyrSerAsnCysHisPheIleLysSerProIleHisLeuGluAsnGlnLys
PheGlyArgArgProGlyGlnSerIleLysIleSerProLysLeuAlaGlnAsnGlyMet
ValGluValIleGlyLeuAspPheLeuSerSerHisTyrHisAlaLeuAlaAlaIleGln
ArgLeuLeuThrAlaThrAsnTyrLysGlyAsnThrLysGlyValValLeuSerArgGlu
SerAsnSerPheGlnPheGluGlyTrpIleProArgIleArgPheThrLysThrGluPhe
LeuGluAlaTyrGlyValLysArgTyrLysThrSerArgAsnLysTyrGluPheSerGly
LysGluAlaGluThrAlaLeuGluAlaLeuTyrHisLeuGlyHisGlnProPheLeuIle
ValAlaThrArgThrArgTrpThrAsnGlyThrGlnIleValAspArgTyrGlnThrLeu
SerProIleIleArgIleTyrGluGlyTrpGluGlyLeuThrAspGluGluAsnIleAsp
IleAspLeuThrProPheAsnSerProProThrArgLysHisLysGlyPheValValGlu
ProCysProIleLeuValAspGlnIleGluSerTyrPheValIleLysProAlaAsnVal
TyrGlnGluIleLysMetArgPheProAsnAlaSerLysTyrAlaTyrThrPheIleAsp
TrpValIleThrAlaAlaAlaLysLysArgArgLysLeuThrLysAspAsnSerTrpPro

GluAsnLeuLeuLeuAsnValAsnValLysSerLeuAlaTyrIleLeuArgMetAsnArg
TyrIleCysThrArgAsnTrpLysLysIleGluLeuAlaIleAspLysCysIleGluIle
AlaIleGlnLeuGlyTrpLeuSerArgArgLysArgIleGluPheLeuAspSerSerLys
LeuSerLysLysGluIleLeuTyrLeuAsnLysGluArgPheGluGluIleThrLysLys
SerLysGluGlnMetGluGlnLeuGluGlnGluSerIleAsnEnd

or parts of it.

14. Protein expressed by the gene according to claim 6 and characterized by the following aminoacid sequence:

pgp3:

MetGlyAsnSerGlyPheTyrLeuTyrAsnThrGluAsnCysValPheAlaAspAsnIle
LysValGlyGlnMetThrGluProLeuLysAspGlnGlnIleIleLeuGlyThrThrSer
ThrProValAlaAlaLysMetThrAlaSerAspGlyIleSerLeuThrValSerAsnAsn
SerSerThrAsnAlaSerIleThrIleGlyLeuAspAlaGluLysAlaTyrGlnLeuIle
LeuGluLysLeuGlyAspGlnIleLeuAspGlyIleAlaAspThrIleValAspSerThr
ValGlnAspIleLeuAspLysIleLysThrAspProSerLeuGlyLeuLeuLysAlaPhe
AsnAsnPheProIleThrAsnLysIleGlnCysAsnGlyLeuPheThrProSerAsnIle
GluThrLeuLeuGlyGlyThrGluIleGlyLysPheThrValThrProLysSerSerGly
SerMetPheLeuValSerAlaAspIleIleAlaSerArgMetGluGlyGlyValValLeu
AlaLeuValArgGluGlyAspSerLysProCysAlaIleSerTyrGlyTyrSerSerGly
IleProAsnLeuCysSerLeuArgThrSerIleThrAsnThrGlyLeuThrProThrThr
TyrSerLeuArgValGlyGlyLeuGluSerGlyValValTrpValAsnAlaLeuSerAsn
GlyAsnAspIleLeuGlyIleThrAsnThrSerAsnValSerPheLeuGluValIlePro
GlnThrAsnAlaEnd

or parts of it.

15. Protein expressed by the gene according to claim 7 and characterized by the following aminoacid sequence:

pgp4:

MetGlnAsnLysArgLysValArgAspAspPheIleLysIleValLysAspValLysLys
AspPheProGluLeuAspLeuLysIleArgValAsnLysGluLysValThrPheLeuAsn
SerProLeuGluLeuTyrHisLysSerValSerLeuIleLeuGlyLeuLeuGlnGlnIle
GluAsnSerLeuGlyLeuPheProAspSerProValLeuGluLysLeuGluAspAsnSer
LeuLysLeuLysLysAlaLeuIleMetLeuIleLeuSerArgLysAspMetPheSerLys
AlaGluEnd

or parts of it.

16. Protein expressed by the gene according to claim 8 and characterized by the following aminoacid sequence:

pgp5:

LeuHisThrLeuValPheCysSerPheLysGlyGlyThrGlyLysThrThrLeuSerLeu
AsnValGlyCysAsnLeuAlaGlnPheLeuGlyLysLysValLeuLeuAlaAspLeuAsp
ProGlnSerAsnLeuSerSerGlyLeuGlyAlaSerValArgSerAspGlnLysGlyLeu
HisAspIleValTyrThrSerAsnAspLeuLysSerIleIleCysGluThrLysLysAsp
SerValAspLeuIleProAlaSerPheSerSerGluGlnPheArgGluLeuAspIleHis
ArgGlyProSerAsnAsnLeuLysLeuPheLeuAsnGluTyrCysAlaProPheTyrAsp
IleCysIleIleAspThrProProSerLeuGlyGlyLeuThrLysGluAlaPheValAla
GlyAspLysLeuIleAlaCysLeuThrProGluProPheSerIleLeuGlyLeuGlnLys
IleArgGluPheLeuSerSerValGlyLysProGluGluGluHisIleLeuGlyIleAla
LeuSerPheTrpAspAspArgAsnSerThrAsnGlnMetTyrIleAspIleIleGluSer
IleTyrLysAsnLysLeuPheSerThrLysIleArgArgAspIleSerLeuSerArgSer
LeuLeuLysGluAspSerValAlaAsnValTyrProAsnSerArgAlaAlaGluAspIle
LeuLysLeuThrHisGluIleAlaAsnIleLeuHisIleGluTyrGluArgAspTyrSer
GlnArgThrThrEnd

or parts of it.

17. Protein expressed by the gene according to claim 9 and characterized by the following aminoacid sequence:

pgp6:

ValAsnLysLeuLysLysGluAlaAspValPhePheLysLysAsnGlnThrAlaAlaSer
LeuAspPheLysLysThrLeuProSerIleGluLeuPheSerAlaThrLeuAsnSerGlu
GluSerGlnSerLeuAspArgLeuPheLeuSerGluSerGlnAsnTyrSerAspGluGlu
PheTyrGlnGluAspIleLeuAlaValLysLeuLeuThrGlyGlnIleLysSerIleGln
LysGlnHisValLeuLeuLeuGlyGluLysIleTyrAsnAlaArgLysIleLeuSerLys
AspHisPheSerSerThrThrPheSerSerTrpIleGluLeuValPheArgThrLysSer
SerAlaTyrAsnAlaLeuAlaTyrTyrGluLeuPheIleAsnLeuProAsnGlnThrLeu
GlnLysGluPheGlnSerIleProTyrLysSerAlaTyrIleLeuAlaAlaArgLysGly
AspLeuLysThrLysValAspValIleGlyLysValCysGlyMetSerAsnSerSerAla
IleArgValLeuAspGlnPheLeuProSerSerArgAsnLysAspValArgGluThrIle
AspLysSerAspSerGluLysAsnArgGlnLeuSerAspPheLeuIleGluIleLeuArg
IleMetCysSerGlyValSerLeuSerSerTyrAsnGluAsnLeuLeuGlnGlnLeuPhe
GluLeuPheLysGlnLysSerEnd

or parts of it.

18. Protein expressed by the gene according to claim 10 and characterized by the following aminoacid sequence:

pgp7 :

LeuValLysLysMetGlySerMetAlaPheHisLysSerArgLeuPheLeuThrPheGly
AspAlaSerGluIleTrpLeuSerThrLeuSerTyrLeuThrArgLysAsnTyrAlaSer
GlyIleAsnPheLeuValSerLeuGluIleLeuAspLeuSerGluThrLeuIleLysAla
IleSerLeuAspHisSerGluSerLeuPheLysIleLysSerLeuAspValPheAsnGly

LysValValSerGluAlaSerLysGlnAlaArgAlaAlaCysTyrIleSerPheThrLys
PheLeuTyrArgLeuThrLysGlyTyrIleLysProAlaIleProLeuLysAspPheGly
AsnThrThrPhePheLysIleArgAspLysIleLysThrGluSerIleSerLysGlnGlu
TrpThrValPhePheGluAlaLeuArgIleValAsnTyrArgAspTyrLeuIleGlyLys
LeuIleValGlnGlyIleArgLysLeuAspGluIleLeuSerLeuArgThrAspAspLeu
PhePheAlaSerAsnGlnIleSerPheArgIleLysLysArgGlnAsnLysGluThrLys
IleLeuIleThrPheProIleSerLeuMetGluGluLeuGlnLysTyrThrCysGlyArg
AsnGlyArgValPheValSerLysIleGlyIleProValThrThrSerGlnValAlaHis
AsnPheArgLeuAlaGluPheHisSerAlaMetLysIleLysIleThrProArgValLeu
ArgAlaSerAlaLeuIleHisLeuLysGlnIleGlyLeuLysAspGluGluIleMetArg
IleSerCysLeuSerSerArgGlnSerValCysSerTyrCysSerGlyGluGluValIle
ProLeuValGlnThrProThrIleLeuEnd

or parts of it.

19. Protein expressed by the gene according to claim 11 and characterized by the following aminoacid sequence:

pgp8 :

MetGlyLysGlyIleLeuSerLeuGlnGlnGluMetSerLeuGluTyrSerGluLysSer
TyrGlnGluValLeuLysIleArgGlnGluSerTyrTrpLysArgMetLysSerPheSer
LeuPheGluValIleMetHisTrpThrAlaSerLeuAsnLysHisThrCysArgSerTyr
ArgGlySerPheLeuSerLeuGluLysIleGlyLeuLeuSerLeuAspMetAsnLeuGln
GluPheSerLeuLeuAsnHisAsnLeuIleLeuAspAlaIleLysLysValSerSerAla
LysThrSerTrpThrGluGlyThrLysGlnValArgAlaAlaSerTyrIleSerLeuThr
ArgPheLeuAsnArgMetThrGlnGlyIleValAlaIleAlaGlnProSerLysGlnGlu
AsnSerArgThrPhePheLysThrArgGluIleValLysThrAspAlaMetAsnSerLeu
GlnThrAlaSerPheLeuLysGluLeuLysLysIleAsnAlaArgAspTrpLeuIleAla
GlnThrMetLeuGlnGlyGlyLysArgSerSerGluValLeuSerLeuGluIleSerGln
IleCysPheGlnGlnAlaThrIleSerPheSerGlnLeuLysAsnArgGlnThrGluLys
ArgIleIleIleThrTyrProGlnLysPheMetHisPheLeuGlnGluTyrIleGlyGln

ArgArgGlyPheValPheValThrArgSerGlyLysMetValGlyLeuArgGlnIleAla

ArgThrPheSerGlnAlaGlyLeuGlnAlaAlaIleProPheLysIleThrProHisVal

LeuArgAlaThrAlaValThrGluTyrLysArgLeuGlyCysSerAspSerAspIleMet

LysValThrGlyHisAlaThrAlaLysMetIlePheAlaTyrAspLysSerSerArgGlu

AspAsnAlaSerLysLysMetAlaLeuIleEnd

or parts of it.

20. Recombinant expression vectors characterized by containing the genes according to claims 4-11.

21. Expression vector according to claim 20 in which the vector pertains to the pEX34 family, the cloned insert is a gene according to claims 4-11, the host cell is E.coli K12ΔH1Δtrp.

22. pO3/GO/MC1 plasmid, constituted by the recombinant expression vector pEX34 and a ORF3D insert.

23. Escherichia coli transformed with the recombinant expression vector according to claim 22 and deposited as ATCC 68315.

24. Process for preparing the immunogenic protein according to claims 12-19 in which:
    a) an ORF is isolated according to claims 4-11
    b) said ORF is cloned in an expression vector and the thus obtained recombinant vector is isolated
    c) bacterial cells are transformed with the aid of a recombinant vector of stage (b)
    d) the bacterial cells transformed as in (c) are cultivated in a suitable medium
    e) the thus obtained protein is isolated and purified from the cell lysate.

25. Process according to claim 24 in which the vector as per stage (b) is pEX34.

26. Process according to claim 25 in which the ORF as per stage (a) is ORF3D.

27. Process according to claim 26 in which the cells as per stage (d) are the ones deposited as ATCC 68315 and the protein product is a recombinant protein (MS2-pgp3) constituted by a terminal portion generated by the vector and by the portion of the pgp3D protein.

28. Process according to claim 27 in which the cell lysate obtained from strain ATCC 68315 is partially purified by dialysis against a phosphate buffer consisting of 0.4% KCl, 0.4% KH_2PO_4, 16% NaCl, 2.5% NaH_2PO_4 at 4°C for about 15 hours, the thus obtained precipitate is discarded and the protein solution is utilized both as such as an antigen in diagnostic tests and further purified.

29. Recombinant MS2-pgp3D protein resulting from the process according to claim 26 and represented by the aminoacid sequence:

-106
MetSerLysThrThrLysLysPheAsnSerLeuCysIleAspLeuProArgAspLeuSer

LeuGluIleTyrGlnSerIleAlaSerValAlaThrGlySerGlyAspProHisSerAsp

AspPheThrAlaIleAlaTyrLeuArgAspGluLeuLeuThrLysHisProThrLeuGly

SerGlyAsnAspGluAlaThrArgArgThrLeuAlaIleAlaLysLeuArgGluAlaAsn

GlyAspArgGlyGlnIleAsnArgGluGlyPheLeuHisAspLysSerLeuSerTrpAsp
+1
IleArgAlaThrGlySerMetGlyAsnSerGlyPheTyrLeuTyrAsnThrGluAsnCys

ValPheAlaAspAsnIleLysValGlyGlnMetThrGluProLeuLysAspGlnGlnIle

IleLeuGlyThrThrSerThrProValAlaAlaLysMetThrAlaSerAspGlyIleSer

LeuThrValSerAsnAsnSerSerThrAsnAlaSerIleThrIleGlyLeuAspAlaGlu

LysAlaTyrGlnLeuIleLeuGluLysLeuGlyAspGlnIleLeuAspGlyIleAlaAsp

ThrIleValAspSerThrValGlnAspIleLeuAspLysIleLysThrAspProSerLeu

GlyLeuLeuLysAlaPheAsnAsnPheProIleThrAsnLysIleGlnCysAsnGlyLeu

PheThrProSerAsnIleGluThrLeuLeuGlyGlyThrGluIleGlyLysPheThrVal

ThrProLysSerSerGlySerMetPheLeuValSerAlaAspIleIleAlaSerArgMet

GluGlyGlyValValLeuAlaLeuValArgGluGlyAspSerLysProCysAlaIleSer

TyrGlyTyrSerSerGlyIleProAsnLeuCysSerLeuArgThrSerIleThrAsnThr

GlyLeuThrProThrThrTyrSerLeuArgValGlyGlyLeuGluSerGlyValValTrp

ValAsnAlaLeuSerAsnGlyAsnAspIleLeuGlyIleThrAsnThrSerAsnValSer

PheLeuGluValIleProGlnThrAsnAlaEnd

or parts thereof.

30. Vaccine against infections caused by Chlamydia trachomatis containing an immunologically effective amount of one of the proteins according to claims 12-19 and 29 and a pharmaceutically acceptable diluent.

31. Vaccine according to claim 30 in which the protein is the one according to claim 14.

32. Vaccine according to claim 30 in which the protein is MS2-pgp3D2.

33. Kit for immunological RIA or ELISA assays in which the antigen utilized in the search for specific antibodies to Chlamydia trachomatis is the protein according to claim 29.

FIG. 1A (1)

```
        10                  30                  50
ATATTCATATTCTGTTGCCAGAAAAAACACCTTTAGGCTATATTAGAGCCATCTTCTTTG

        70                  90                 110
AAGCGTTGTCTTCTCGAGAAGATTTATCGTACGCAAATATCATCTTTGCGGTTGCGTGTC

       130                 150                 170
CTGTGACCTTCATTATGTCGGAGTCTGAGCACCCTAGGCGTTTGTACTCCGTCACAGCGG

       190                 210                 230
TTGCTCGAAGCACGTGCGGGGTTATTTTAAAAGGGATTGCAGCTTGTAGTCCTGCTTGAG

       250                 270                 290
AGAACGTGCGGGCGATTTGCCTTAACCCCACCATTTTTCCGGAGCGAGTTACGAAGACAA

       310                 330                 350
AACCTCTTCGTTGACCGATGTACTCTTGTAGAAAGTGCATAAACTTCTGAGGATAAGTTA

       370                 390                 410
TAATAATCCTCTTTTCTGTCTGACGGTTCTTAAGCTGGGAGAAAGAAATGGTAGCTTGTT

       430                 450                 470
GGAAACAAATCTGACTAATCTCCAAGCTTAAGACTTCAGAGGAGCGTTTACCTCCTTGGA

       490                 510                 530
GCATTGTCTGGGCGATCAACCAATCCCGGGCATTGATTTTTTTTAGCTCTTTTAGGAAGG

       550                 570                 590
ATGCTGTTTGCAAACTGTTCATCGCATCCGTTTTTACTATTTCCCTGGTTTTAAAAAATG

       610                 630                 650
TTCGACTATTTTCTTGTTTAGAAGGTTGCGCTATAGCGACTATTCCTTGAGTCATCCTGT

       670                 690                 710
TTAGGAATCTTGTTAAGGAAATATAGCTTGCTGCTCGAACTTGTTTAGTACCTTCGGTCC

       730                 750                 770
AAGAAGTCTTGGCAGAGGAAACTTTTTTAATCGCATCTAGGATTAGATTATGATTTAAAA

       790                 810                 830
GGGAAAACTCTTGCAGATTCATATCCAAGGACAATAGACCAATCTTTTCTAAAGACAAAA

       850                 870                 890
AAGATCCTCGATATGATCTACAAGTATGTTTGTTGAGTGATGCGGTCCAATGCATAATAA

       910                 930                 950
CTTCGAATAAGGAGAAGCTTTTCATGCGTTTCCAATAGGATTCTTGGCGAATTTTTAAAA

       970                 990                1010
CTTCCTGATAAGACTTTTCACTATATTCTAACGACATTTCTTGCTGCAAAGATAAAATCC

      1030                1050                1070
CTTTACCCATGAAATCCCTCGTGATATAACCTATCCGTAAAATGTCCTGATTAGTGAAAT

      1090                1110                1130
AATCAGGTTGTTAACAGGATAGCACGCTCGGTATTTTTTTATATAAACATGAAAACTCGT
                                          ORF1 >>  MetLysThrArg
```

FIG. 1A (2)

```
      1150                1170                1190
TCCGAAATAGAAAATCGCATGCAAGATATCGAGTATGCGTTGTTAGGTAAAGCTCTGATA
SerGluIleGluAsnArgMetGlnAspIleGluTyrAlaLeuLeuGlyLysAlaLeuIle

      1210                1230                1250
TTTGAAGACTCTACTGAGTATATTCTGAGGCAGCTTGCTAATTATGAGTTTAAGTGTTCT
PheGluAspSerThrGluTyrIleLeuArgGlnLeuAlaAsnTyrGluPheLysCysSer

      1270                1290                1310
CATCATAAAAACATATTCATAGTATTTAAACACTTAAAAGACAATGGATTACCTATAACT
HisHisLysAsnIlePheIleValPheLysHisLeuLysAspAsnGlyLeuProIleThr

      1330                1350                1370
GTAGACTCGGCTTGGGAAGAGCTTTTGCGGCGTCGTATCAAAGATATGGACAAATCGTAT
ValAspSerAlaTrpGluGluLeuLeuArgArgArgIleLysAspMetAspLysSerTyr

      1390                1410                1430
CTCGGGTTAATGTTGCATGATGCTTTATCAAATGACAAGCTTAGATCCGTTTCTCATACG
LeuGlyLeuMetLeuHisAspAlaLeuSerAsnAspLysLeuArgSerValSerHisThr

      1450                1470                1490
GTTTTCCTCGATGATTTGAGCGTGTGTAGCGCTGAAGAAAATTTGAGTAATTTCATTTTC
ValPheLeuAspAspLeuSerValCysSerAlaGluGluAsnLeuSerAsnPheIlePhe

      1510                1530                1550
CGCTCGTTTAATGAGTACAATGAAAATCCATTGCGTAGATCTCCGTTTCTATTGCTTGAG
ArgSerPheAsnGluTyrAsnGluAsnProLeuArgArgSerProPheLeuLeuLeuGlu

      1570                1590                1610
CGTATAAAGGGAAGGCTTGATAGTGCTATAGCAAAGACTTTTTCTATTCGCAGCGCTAGA
ArgIleLysGlyArgLeuAspSerAlaIleAlaLysThrPheSerIleArgSerAlaArg

      1630                1650                1670
GGCCGGTCTATTTATGATATATTCTCACAGTCAGAAATTGGAGTGCTGGCTCGTATAAAA
GlyArgSerIleTyrAspIlePheSerGlnSerGluIleGlyValLeuAlaArgIleLys

      1690                1710                1730
AAAAGACGAGTAGCGTTCTCTGAGAATCAAAATTCTTTCTTTGATGGCTTCCCAACAGGA
LysArgArgValAlaPheSerGluAsnGlnAsnSerPhePheAspGlyPheProThrGly

      1750                1770                1790
TACAAGGATATTGATGATAAAGGAGTTATCTTAGCTAAAGGTAATTTCGTGATTATAGCA
TyrLysAspIleAspAspLysGlyValIleLeuAlaLysGlyAsnPheValIleIleAla

      1810                1830                1850
GCTAGACCATCTATAGGGAAAACAGCTTTAGCTATAGACATGGCGATAAATCTTGCGGTT
AlaArgProSerIleGlyLysThrAlaLeuAlaIleAspMetAlaIleAsnLeuAlaVal

      1870                1890                1910
ACTCAACAGCGTAGAGTTGGTTTCCTATCTCTAGAAATGAGCGCAGGTCAAATTGTTGAG
ThrGlnGlnArgArgValGlyPheLeuSerLeuGluMetSerAlaGlyGlnIleValGlu

      1930                1950                1970
CGGATTATTGCTAATTTAACAGGAATATCTGGTGAAAAATTACAAAGAGGGGATCTCTCT
ArgIleIleAlaAsnLeuThrGlyIleSerGlyGluLysLeuGlnArgGlyAspLeuSer
```

FIG. 1A (3)

```
      1990                2010                2030
AAAGAAGAATTATTCCGAGTAGAAGAAGCTGGAGAAACGGTTAGAGAATCACATTTTTAT
LysGluGluLeuPheArgValGluGluAlaGlyGluThrValArgGluSerHisPheTyr

      2050                2070                2090
ATCTGCAGTGATAGTCAGTATAAGCTTAACTTAATCGCGAATCAGATCCGGTTGCTGAGA
IleCysSerAspSerGlnTyrLysLeuAsnLeuIleAlaAsnGlnIleArgLeuLeuArg

      2110                2130                2150
AAAGAAGATCGAGTAGACGTAATATTTATCGATTACTTGCAGTTGATCAACTCATCGGTT
LysGluAspArgValAspValIlePheIleAspTyrLeuGlnLeuIleAsnSerSerVal

      2170                2190                2210
GGAGAAAATCGTCAAAATGAAATAGCAGATATATCTAGAACCTTAAGAGGTTTAGCCTCA
GlyGluAsnArgGlnAsnGluIleAlaAspIleSerArgThrLeuArgGlyLeuAlaSer

      2230                2250                2270
GAGCTAAACATTCCTATAGTTTGTTTATCCCAACTATCTAGAAAAGTTGAGGATAGAGCA
GluLeuAsnIleProIleValCysLeuSerGlnLeuSerArgLysValGluAspArgAla

      2290                2310                2330
AATAAAGTTCCCATGCTTTCAGATTTGCGAGACAGCGGTCAAATAGAGCAAGACGCAGAT
AsnLysValProMetLeuSerAspLeuArgAspSerGlyGlnIleGluGlnAspAlaAsp

      2350                2370                2390
GTGATTTTGTTTATCAATAGGAAGGAATCGTCTTCTAATTGTGAGATAACTGTTGGGAAA
ValIleLeuPheIleAsnArgLysGluSerSerSerAsnCysGluIleThrValGlyLys

      2410                2430                2450
AATAGACATGGATCGGTTTTCTCTTCGGTATTACATTTCGATCCAAAAATTAGTAAATTC
AsnArgHisGlySerValPheSerSerValLeuHisPheAspProLysIleSerLysPhe

      2470                2490                2510
TCCGCTATTAAAAAAGTATGGTAAATTATAGTAACTGCCACTTCATCAAAAGTCCTATCC
SerAlaIleLysLysValTrpEnd
          ORF2 >> MetValAsnTyrSerAsnCysHisPheIleLysSerProIleH

      2530                2550                2570
ACCTTGAAAATCAGAAGTTTGGAAGAAGACCTGGTCAATCTATTAAGATATCTCCCAAAT
isLeuGluAsnGlnLysPheGlyArgArgProGlyGlnSerIleLysIleSerProLysL

      2590                2610                2630
TGGCTCAAAATGGGATGGTAGAAGTTATAGGTCTTGATTTTCTTTCATCTCATTACCATG
euAlaGlnAsnGlyMetValGluValIleGlyLeuAspPheLeuSerSerHisTyrHisA

      2650                2670                2690
CATTAGCAGCTATCCAAAGATTACTGACCGCAACGAATTACAAGGGGAACACAAAAGGGG
laLeuAlaAlaIleGlnArgLeuLeuThrAlaThrAsnTyrLysGlyAsnThrLysGlyV

      2710                2730                2750
TTGTTTTATCCAGAGAATCAAATAGTTTTCAATTTGAAGGATGGATACCAAGAATCCGTT
alValLeuSerArgGluSerAsnSerPheGlnPheGluGlyTrpIleProArgIleArgP

      2770                2790                2810
TTACAAAAACTGAATTCTTAGAGGCTTATGGAGTTAAGCGGTATAAAACATCCAGAAATA
heThrLysThrGluPheLeuGluAlaTyrGlyValLysArgTyrLysThrSerArgAsnL
```

FIG. 1A (4)

```
      2830                2850                2870
AGTATGAGTTTAGTGGAAAAGAAGCTGAAACTGCTTTAGAAGCCTTATACCATTTAGGAC
ysTyrGluPheSerGlyLysGluAlaGluThrAlaLeuGluAlaLeuTyrHisLeuGlyH

      2890                2910                2930
ATCAACCGTTTTTAATAGTGGCAACTAGAACTCGATGGACTAATGGAACACAAATAGTAG
isGlnProPheLeuIleValAlaThrArgThrArgTrpThrAsnGlyThrGlnIleValA

      2950                2970                2990
ACCGTTACCAAACTCTTTCTCCGATCATTAGGATTTACGAAGGATGGGAAGGTTTAACTG
spArgTyrGlnThrLeuSerProIleIleArgIleTyrGluGlyTrpGluGlyLeuThrA

      3010                3030                3050
ACGAAGAAAATATAGATATAGACTTAACACCTTTTAATTCACCACCTACACGGAAACATA
spGluGluAsnIleAspIleAspLeuThrProPheAsnSerProProThrArgLysHisL

      3070                3090                3110
AAGGGTTCGTTGTAGAGCCATGTCCTATCTTGGTAGATCAAATAGAATCCTACTTTGTAA
ysGlyPheValValGluProCysProIleLeuValAspGlnIleGluSerTyrPheValI

      3130                3150                3170
TCAAGCCTGCAAATGTATACCAAGAAATAAAAATGCGTTTCCCAAATGCATCAAAGTATG
leLysProAlaAsnValTyrGlnGluIleLysMetArgPheProAsnAlaSerLysTyrA

      3190                3210                3230
CTTACACATTTATCGACTGGGTGATTACAGCAGCTGCGAAAAAGAGACGAAAATTAACTA
laTyrThrPheIleAspTrpValIleThrAlaAlaAlaLysLysArgArgLysLeuThrL

      3250                3270                3290
AGGATAATTCTTGGCCAGAAAACTTGTTATTAAACGTTAACGTTAAAAGTCTTGCATATA
ysAspAsnSerTrpProGluAsnLeuLeuLeuAsnValAsnValLysSerLeuAlaTyrI

      3310                3330                3350
TTTTAAGGATGAATCGGTACATCTGTACAAGGAACTGGAAAAAAATCGAGTTAGCTATCG
leLeuArgMetAsnArgTyrIleCysThrArgAsnTrpLysLysIleGluLeuAlaIleA

      3370                3390                3410
ATAAATGTATAGAAATCGCCATTCAGCTTGGCTGGTTATCTAGAAGAAAACGCATTGAAT
spLysCysIleGluIleAlaIleGlnLeuGlyTrpLeuSerArgArgLysArgIleGluP

      3430                3450                3470
TTCTGGATTCTTCTAAACTCTCTAAAAAAGAAATTCTATATCTAAATAAAGAGCGCTTTG
heLeuAspSerSerLysLeuSerLysLysGluIleLeuTyrLeuAsnLysGluArgPheG

      3490                3510                3530
AAGAAATAACTAAGAAATCTAAAGAACAAATGGAACAATTAGAACAAGAATCTATTAATT
luGluIleThrLysLysSerLysGluGlnMetGluGlnLeuGluGlnGluSerIleAsnE

      3550                3570                3590
AATAGCAAGCTTGAAACTAAAAACCTAATTTATTTAAAGCTCAAAATAAAAAAAGAGTTTT
nd

      3610                3630                3650
AAAATGGGAAATTCTGGTTTTTATTTGTATAACACTGAAAACTGCGTCTTTGCTGATAAT
ORF3>> MetGlyAsnSerGlyPheTyrLeuTyrAsnThrGluAsnCysValPheAlaAspAsn

      3670                3690                3710
ATCAAAGTTGGGCAAATGACAGAGCCGCTCAAGGACCAGCAAATAATCCTTGGGACAACA
IleLysValGlyGlnMetThrGluProLeuLysAspGlnGlnIleIleLeuGlyThrThr
```

FIG. 14 (5)

```
      3730                3750                3770
TCAACACCTGTCGCAGCCAAAATGACAGCTTCTGATGGAATATCTTTAACAGTCTCCAAT
SerThrProValAlaAlaLysMetThrAlaSerAspGlyIleSerLeuThrValSerAsn

      3790                3810                3830
AATTCATCAACCAATGCTTCTATTACAATTGGTTTGGATGCGGAAAAAGCTTACCAGCTT
AsnSerSerThrAsnAlaSerIleThrIleGlyLeuAspAlaGluLysAlaTyrGlnLeu

      3850                3870                3890
ATTCTAGAAAAGTTGGGAGATCAAATTCTTGATGGAATTGCTGATACTATTGTTGATAGT
IleLeuGluLysLeuGlyAspGlnIleLeuAspGlyIleAlaAspThrIleValAspSer

      3910                3930                3950
ACAGTCCAAGATATTTTAGACAAAATCAAAACAGACCCTTCTCTAGGTTTGTTGAAAGCT
ThrValGlnAspIleLeuAspLysIleLysThrAspProSerLeuGlyLeuLeuLysAla

      3970                3990                4010
TTTAACAACTTTCCAATCACTAATAAAATTCAATGCAACGGGTTATTCACTCCCAGTAAC
PheAsnAsnPheProIleThrAsnLysIleGlnCysAsnGlyLeuPheThrProSerAsn

      4030                4050                4070
ATTGAAACTTTATTAGGAGGAACTGAAATAGGAAAATTCACAGTCACACCCAAAAGCTCT
IleGluThrLeuLeuGlyGlyThrGluIleGlyLysPheThrValThrProLysSerSer

      4090                4110                4130
GGGAGCATGTTCTTAGTCTCAGCAGATATTATTGCATCAAGAATGGAAGGCGGCGTTGTT
GlySerMetPheLeuValSerAlaAspIleIleAlaSerArgMetGluGlyGlyValVal

      4150                4170                4190
CTAGCTTTGGTACGAGAAGGTGATTCTAAGCCCTGCGCGATTAGTTATGGATACTCATCA
LeuAlaLeuValArgGluGlyAspSerLysProCysAlaIleSerTyrGlyTyrSerSer

      4210                4230                4250
GGCATTCCTAATTTATGTAGTCTAAGAACCAGTATTACTAATACAGGATTGACTCCGACA
GlyIleProAsnLeuCysSerLeuArgThrSerIleThrAsnThrGlyLeuThrProThr

      4270                4290                4310
ACGTATTCATTACGTGTAGGCGGTTTAGAAAGCGGTGTGGTATGGGTTAATGCCCTTTCT
ThrTyrSerLeuArgValGlyGlyLeuGluSerGlyValValTrpValAsnAlaLeuSer

      4330                4350                4370
AATGGCAATGATATTTTAGGAATAACAAATACTTCTAATGTATCTTTTTTAGAGGTAATA
AsnGlyAsnAspIleLeuGlyIleThrAsnThrSerAsnValSerPheLeuGluValIle

      4390                4410                4430
CCTCAAACAAACGCTTAAACAATTTTTATTGGATTTTTCTTATAGGTTTTATATTTAGAG
ProGlnThrAsnAlaEnd

      4450                4470                4490
AAAACAGTTCGAATTACGGGGTTTGTTATGCAAAATAAAAGAAAAGTGAGGGACGATTTT
                  ORF4 >> MetGlnAsnLysArgLysValArgAspAspPhe

      4510                4530                4550
ATTAAAATTGTTAAAGATGTGAAAAAAGATTTCCCCGAATTAGACCTAAAAATACGAGTA
IleLysIleValLysAspValLysLysAspPheProGluLeuAspLeuLysIleArgVal

      4570                4590                4610
AACAAGGAAAAAGTAACTTTCTTAAATTCTCCCTTAGAACTCTACCATAAAAGTGTCTCA
AsnLysGluLysValThrPheLeuAsnSerProLeuGluLeuTyrHisLysSerValSer
```

FIG. 1A (6)

```
      4630                    4650                    4670
CTAATTCTAGGACTGCTTCAACAAATAGAAAACTCTTTAGGATTATTCCCAGACTCTCCT
LeuIleLeuGlyLeuLeuGlnGlnIleGluAsnSerLeuGlyLeuPheProAspSerPro

      4690                    4710                    4730
GTTCTTGAAAAATTAGAGGATAACAGTTTAAAGCTAAAAAAGGCTTTGATTATGCTTATC
ValLeuGluLysLeuGluAspAsnSerLeuLysLeuLysLysAlaLeuIleMetLeuIle

      4750                    4770                    4790
TTGTCTAGAAAAGACATGTTTTCCAAGGCTGAATAGACAACTTACTCTAACGTTGGAGTT
LeuSerArgLysAspMetPheSerLysAlaGluEnd

      4810                    4830                    4850
GATTTGCACACCTTAGTTTTTTGCTCTTTTAAGGGAGGAACTGGAAAAACAACACTTTCT
ORF5 >> LeuHisThrLeuValPheCysSerPheLysGlyGlyThrGlyLysThrThrLeuSer
      4870                    4890                    4910
CTAAACGTGGGATGCAACTTGGCCCAATTTTTAGGGAAAAAAGTGTTACTTGCTGACCTA
LeuAsnValGlyCysAsnLeuAlaGlnPheLeuGlyLysLysValLeuLeuAlaAspLeu

      4930                    4950                    4970
GACCCGCAATCCAATTTATCTTCTGGATTGGGGGCTAGTGTCAGAAGTGACCAAAAAGGC
AspProGlnSerAsnLeuSerSerGlyLeuGlyAlaSerValArgSerAspGlnLysGly

      4990                    5010                    5030
TTGCACGACATAGTATACACATCAAACGATTTAAAATCAATCATTTGCGAAACAAAAAAA
LeuHisAspIleValTyrThrSerAsnAspLeuLysSerIleIleCysGluThrLysLys

      5050                    5070                    5090
GATAGTGTGGACCTAATTCCTGCATCATTTTCATCCGAACAGTTTAGAGAATTGGATATT
AspSerValAspLeuIleProAlaSerPheSerSerGluGlnPheArgGluLeuAspIle

      5110                    5130                    5150
CATAGAGGACCTAGTAACAACTTAAAGTTATTTCTGAATGAGTACTGCGCTCCTTTTTAT
HisArgGlyProSerAsnAsnLeuLysLeuPheLeuAsnGluTyrCysAlaProPheTyr

      5170                    5190                    5210
GACATCTGCATAATAGACACTCCACCTAGCCTAGGAGGGTTAACGAAAGAAGCTTTTGTT
AspIleCysIleIleAspThrProProSerLeuGlyGlyLeuThrLysGluAlaPheVal

      5230                    5250                    5270
GCAGGAGACAAATTAATTGCTTGTTTAACTCCAGAACCTTTTTCTATTCTAGGGTTACAA
AlaGlyAspLysLeuIleAlaCysLeuThrProGluProPheSerIleLeuGlyLeuGln

      5290                    5310                    5330
AAGATACGTGAATTCTTAAGTTCGGTCGGAAAACCTGAAGAAGAACACATTCTTGGAATA
LysIleArgGluPheLeuSerSerValGlyLysProGluGluGluHisIleLeuGlyIle

      5350                    5370                    5390
GCTTTGTCTTTTTGGGATGATCGTAACTCGACTAACCAAATGTATATAGACATTATCGAG
AlaLeuSerPheTrpAspAspArgAsnSerThrAsnGlnMetTyrIleAspIleIleGlu

      5410                    5430                    5450
TCTATTTACAAAAACAAGCTTTTTTCAACAAAAATTCGTCGAGATATTTCTCTCAGCCGT
SerIleTyrLysAsnLysLeuPheSerThrLysIleArgArgAspIleSerLeuSerArg

      5470                    5490                    5510
TCTCTTCTTAAAGAAGATTCTGTAGCTAATGTCTATCCAAATTCTAGGGCCGCAGAAGAT
SerLeuLeuLysGluAspSerValAlaAsnValTyrProAsnSerArgAlaAlaGluAsp
```

FIG. 1A (7)

```
      5530                    5550                    5570
ATTCTGAAGTTAACGCATGAAATAGCAAATATTTTGCATATCGAATATGAACGAGATTAC
IleLeuLysLeuThrHisGluIleAlaAsnIleLeuHisIleGluTyrGluArgAspTyr

      5590                    5610                    5630
TCTCAGAGGACAACGTGAACAAACTAAAAAAAAGAAGCGGATGTCTTTTTTAAAAAAAATC
SerGlnArgThrThrEnd
      ORF6 >> ValAsnLysLeuLysLysGluAlaAspValPhePheLysLysAsnG

      5650                    5670                    5690
AAACTGCCGCTTCTCTAGATTTTAAGAAGACGCTTCCCTCCATTGAACTATTCTCAGCAA
lnThrAlaAlaSerLeuAspPheLysLysThrLeuProSerIleGluLeuPheSerAlaT

      5710                    5730                    5750
CTTTGAATTCTGAGGAAAGTCAGAGTTTGGATCGATTATTTTTATCAGAGTCCCAAAACT
hrLeuAsnSerGluGluSerGlnSerLeuAspArgLeuPheLeuSerGluSerGlnAsnT

      5770                    5790                    5810
ATTCGGATGAAGAATTTTATCAAGAAGACATCCTAGCGGTAAAACTGCTTACTGGTCAGA
yrSerAspGluGluPheTyrGlnGluAspIleLeuAlaValLysLeuLeuThrGlyGlnI

      5830                    5850                    5870
TAAAATCCATACAGAAGCAACACGTACTTCTTTTAGGAGAAAAAATCTATAATGCTAGAA
leLysSerIleGlnLysGlnHisValLeuLeuLeuGlyGluLysIleTyrAsnAlaArgL

      5890                    5910                    5930
AAATCCTGAGTAAGGATCACTTCTCCTCAACAACTTTTTCATCTTGGATAGAGTTAGTTT
ysIleLeuSerLysAspHisPheSerSerThrThrPheSerSerTrpIleGluLeuValP

      5950                    5970                    5990
TTAGAACTAAGTCTTCTGCTTACAATGCTCTTGCATATTACGAGCTTTTTATAAACCTCC
heArgThrLysSerSerAlaTyrAsnAlaLeuAlaTyrTyrGluLeuPheIleAsnLeuP

      6010                    6030                    6050
CCAACCAAACTCTACAAAAAGAGTTTCAATCGATCCCCTATAAATCCGCATATATTTTGG
roAsnGlnThrLeuGlnLysGluPheGlnSerIleProTyrLysSerAlaTyrIleLeuA

      6070                    6090                    6110
CCGCTAGAAAAGGCGATTTAAAAACCAAGGTCGATGTGATAGGGAAAGTATGTGGAATGT
laAlaArgLysGlyAspLeuLysThrLysValAspValIleGlyLysValCysGlyMetS

      6130                    6150                    6170
CGAACTCATCGGCGATAAGGGTGTTGGATCAATTTCTTCCTTCATCTAGAAACAAAGACG
erAsnSerSerAlaIleArgValLeuAspGlnPheLeuProSerSerArgAsnLysAspV

      6190                    6210                    6230
TTAGAGAAACGATAGATAAGTCTGATTCAGAGAAGAATCGCCAATTATCTGATTTCTTAA
alArgGluThrIleAspLysSerAspSerGluLysAsnArgGlnLeuSerAspPheLeuI

      6250                    6270                    6290
TAGAGATACTTCGCATCATGTGTTCCGGAGTTTCTTTGTCCTCCTATAACGAAAATCTTC
leGluIleLeuArgIleMetCysSerGlyValSerLeuSerSerTyrAsnGluAsnLeuL

      6310                    6330                    6350
TACAACAGCTTTTTGAACTTTTTAAGCAAAAGAGCTGATCCTCCGTCAGCTCATATATAT
euGlnGlnLeuPheGluLeuPheLysGlnLysSerEnd
```

FIG. 1A (8)

```
      6370                6390                6410
ATATCTATTATATATATATATTTAGGGATTTGATTTCACGAGAGAGATTTGCAACTCTTG

      6430                6450                6470
GTGGTAGACTTTGCAACTCTTGGTGGTAGACTTTGCAACTCTTGGTGGTAGACTTTGCAA

      6490                6510                6530
CTCTTGGTGGTAGACTTGGTCATAATGGACTTTTGTTAAAAAATTTATTAAAATCTTAGA

      6550                6570                6590
GCTCCGATTTTGAATAGCTTTGGTTAAGAAAATGGGCTCGATGGCTTTCCATAAAAGTAG
           ORF7 >> LeuValLysLysMetGlySerMetAlaPheHisLysSerAr

      6610                6630                6650
ATTGTTTTTAACTTTTGGGGACGCGTCGGAAATTTGGTTATCTACTTTATCTTATCTAAC
gLeuPheLeuThrPheGlyAspAlaSerGluIleTrpLeuSerThrLeuSerTyrLeuTh

      6670                6690                6710
TAGAAAAAATTATGCGTCTGGGATTAACTTTCTTGTTTCTTTAGAGATTCTGGATTTATC
rArgLysAsnTyrAlaSerGlyIleAsnPheLeuValSerLeuGluIleLeuAspLeuSe

      6730                6750                6770
GGAAACCTTGATAAAGGCTATTTCTCTTGACCACAGCGAATCTTTGTTTAAAATCAAGTC
rGluThrLeuIleLysAlaIleSerLeuAspHisSerGluSerLeuPheLysIleLysSe

      6790                6810                6830
TCTAGATGTTTTTAATGGAAAAGTTGTTTCAGAGGCATCTAAACAGGCTAGAGCGGCATG
rLeuAspValPheAsnGlyLysValValSerGluAlaSerLysGlnAlaArgAlaAlaCy

      6850                6870                6890
CTACATATCTTTCACAAAGTTTTTGTATAGATTGACCAAGGGATATATTAAACCCGCTAT
sTyrIleSerPheThrLysPheLeuTyrArgLeuThrLysGlyTyrIleLysProAlaIl

      6910                6930                6950
TCCATTGAAAGATTTTGGAAACACTACATTTTTTAAAATCCGAGACAAAATCAAAACAGA
eProLeuLysAspPheGlyAsnThrThrPhePheLysIleArgAspLysIleLysThrGl

      6970                6990                7010
ATCGATTTCTAAGCAGGAATGGACAGTTTTTTTTGAAGCGCTCCGGATAGTGAATTATAG
uSerIleSerLysGlnGluTrpThrValPhePheGluAlaLeuArgIleValAsnTyrAr

      7030                7050                7070
AGACTATTTAATCGGTAAATTGATTGTACAAGGGATCCGTAAGTTAGACGAAATTTTGTC
gAspTyrLeuIleGlyLysLeuIleValGlnGlyIleArgLysLeuAspGluIleLeuSe

      7090                7110                7130
TTTGCGCACAGACGATCTATTTTTTGCATCCAATCAGATTTCCTTTCGCATTAAAAAAAG
rLeuArgThrAspAspLeuPhePheAlaSerAsnGlnIleSerPheArgIleLysLysAr

      7150                7170                7190
ACAGAATAAAGAAACCAAAATTCTAATCACATTTCCTATCAGCTTAATGGAAGAGTTGCA
gGlnAsnLysGluThrLysIleLeuIleThrPheProIleSerLeuMetGluGluLeuGl

      7210                7230                7250
AAAATACACTTGTGGGAGAAATGGGAGAGTATTTGTTTCTAAAATAGGGATTCCTGTAAC
nLysTyrThrCysGlyArgAsnGlyArgValPheValSerLysIleGlyIleProValTh
```

FIG. 1A (9)

```
      7270                7290                7310
AACAAGTCAGGTTGCGCATAATTTTAGGCTTGCAGAGTTCCATAGTGCTATGAAAATAAA
rThrSerGlnValAlaHisAsnPheArgLeuAlaGluPheHisSerAlaMetLysIleLy

      7330                7350                7370
AATTACTCCCAGAGTACTTCGTGCAAGCGCTTTGATTCATTTAAAGCAAATAGGATTAAA
sIleThrProArgValLeuArgAlaSerAlaLeuIleHisLeuLysGlnIleGlyLeuLy

      7390                7410                7430
AGATGAGGAAATCATGCGTATTTCCTGTCTTTCATCGAGACAAAGTGTGTGTTCTTATTG
sAspGluGluIleMetArgIleSerCysLeuSerSerArgGlnSerValCysSerTyrCy

      7450                7470                7490
TTCTGGGGAAGAGGTAATTCCTCTAGTACAAACACCCACAATATTGTGATATAATTAAAA
sSerGlyGluGluValIleProLeuValGlnThrProThrIleLeuEnd

TT
```

FIG. 1B (1)

```
GCATGCGATTTTCTATTTCGGAACGAGTTTTCATGTTTATATAAAAAAATACCGAGCGTG

CTATCCTGTTAACAACCTGATTATTTCACTAATCAGGACATTTTACGGATAGGTTATATC

ACGAGGGATTTCATGGGTAAAGGGATTTTATCTTTGCAGCAAGAAATGTCGTTAGAATAT
    ORF8 >> MetGlyLysGlyIleLeuSerLeuGlnGlnGluMetSerLeuGluTyr

AGTGAAAAGTCTTATCAGGAAGTTTTAAAAATTCGCCAAGAATCCTATTGGAAACGCATG
SerGluLysSerTyrGlnGluValLeuLysIleArgGlnGluSerTyrTrpLysArgMet

AAAAGCTTCTCCTTATTCGAAGTTATTATGCATTGGACCGCATCACTCAACAAACATACT
LysSerPheSerLeuPheGluValIleMetHisTrpThrAlaSerLeuAsnLysHisThr

TGTAGATCATATCGAGGATCTTTTTTGTCTTTAGAAAAGATTGGTCTATTGTCCTTGGAT
CysArgSerTyrArgGlySerPheLeuSerLeuGluLysIleGlyLeuLeuSerLeuAsp

ATGAATCTGCAAGAGTTTTCCCTTTTAAATCATAATCTAATCCTAGATGCGATTAAAAAA
MetAsnLeuGlnGluPheSerLeuLeuAsnHisAsnLeuIleLeuAspAlaIleLysLys

GTTTCCTCTGCCAAGACTTCTTGGACCGAAGGTACTAAACAAGTTCGAGCAGCAAGCTAT
ValSerSerAlaLysThrSerTrpThrGluGlyThrLysGlnValArgAlaAlaSerTyr

ATTTCCTTAACAAGATTCCTAAACAGGATGACTCAAGGAATAGTCGCTATAGCGCAACCT
IleSerLeuThrArgPheLeuAsnArgMetThrGlnGlyIleValAlaIleAlaGlnPro

TCTAAACAAGAAAATAGTCGAACATTTTTTAAAACCAGGGAAATAGTAAAAACGGATGCG
SerLysGlnGluAsnSerArgThrPhePheLysThrArgGluIleValLysThrAspAla

ATGAACAGTTTGCAAACAGCATCCTTCCTAAAAGAGCTAAAAAAAATCAATGCCCGGGAT
MetAsnSerLeuGlnThrAlaSerPheLeuLysGluLeuLysLysIleAsnAlaArgAsp

TGGTTGATCGCCCAGACAATGCTCCAAGGAGGTAAACGCTCCTCTGAAGTCTTAAGCTTG
TrpLeuIleAlaGlnThrMetLeuGlnGlyGlyLysArgSerSerGluValLeuSerLeu

GAGATTAGTCAGATTTGTTTCCAACAAGCTACCATTTCTTTCTCCCAGCTTAAGAACCGT
GluIleSerGlnIleCysPheGlnGlnAlaThrIleSerPheSerGlnLeuLysAsnArg

CAGACAGAAAAGAGGATTATTATAACTTATCCTCAGAAGTTTATGCACTTTCTACAAGAG
GlnThrGluLysArgIleIleIleThrTyrProGlnLysPheMetHisPheLeuGlnGlu
```

FIG. 1B (2)

```
TACATCGGTCAACGAAGAGGTTTTGTCTTCGTAACTCGCTCCGGAAAAATGGTGGGGTTA
TyrIleGlyGlnArgArgGlyPheValPheValThrArgSerGlyLysMetValGlyLeu

AGGCAAATCGCCCGCACGTTCTCTCAAGCAGGACTACAAGCTGCAATCCCTTTTAAAATA
ArgGlnIleAlaArgThrPheSerGlnAlaGlyLeuGlnAlaAlaIleProPheLysIle

ACCCCGCACGTGCTTCGAGCAACCGCTGTGACGGAGTACAAACGCCTAGGGTGCTCAGAC
ThrProHisValLeuArgAlaThrAlaValThrGluTyrLysArgLeuGlyCysSerAsp

TCCGACATAATGAAGGTCACAGGACACGCAACCGCAAAGATGATATTTGCGTACGATAAA
SerAspIleMetLysValThrGlyHisAlaThrAlaLysMetIlePheAlaTyrAspLys

TCTTCTCGAGAAGACAACGCTTCAAAGAAGATGGCTCTAATATAGCCTAAAGGTGTTTTT
SerSerArgGluAspAsnAlaSerLysLysMetAlaLeuIleEnd

TCTGGCAACAGAATATGAATAT
```

FIG. 2

```
      3610                3630                3650
   AAAATGGGAAATTCTGGTTTTTATTTGTATAACACTGAAAACTGCGTCTTTGCTGATAAT
ORF3>> MetGlyAsnSerGlyPheTyrLeuTyrAsnThrGluAsnCysValPheAlaAspAsn

      3670                3690                3710
   ATCAAAGTTGGGCAAATGACAGAGCCGCTCAAGGACCAGCAAATAATCCTTGGGACAACA
   IleLysValGlyGlnMetThrGluProLeuLysAspGlnGlnIleIleLeuGlyThrThr

      3730                3750                3770
   TCAACACCTGTCGCAGCCAAAATGACAGCTTCTGATGGAATATCTTTAACAGTCTCCAAT
   SerThrProValAlaAlaLysMetThrAlaSerAspGlyIleSerLeuThrValSerAsn

      3790                3810                3830
   AATTCATCAACCAATGCTTCTATTACAATTGGTTTGGATGCGGAAAAAGCTTACCAGCTT
   AsnSerSerThrAsnAlaSerIleThrIleGlyLeuAspAlaGluLysAlaTyrGlnLeu

      3850                3870                3890
   ATTCTAGAAAAGTTGGGAGATCAAATTCTTGATGGAATTGCTGATACTATTGTTGATAGT
   IleLeuGluLysLeuGlyAspGlnIleLeuAspGlyIleAlaAspThrIleValAspSer

      3910                3930                3950
   ACAGTCCAAGATATTTTAGACAAAATCAAAACAGACCCTTCTCTAGGTTTGTTGAAAGCT
   ThrValGlnAspIleLeuAspLysIleLysThrAspProSerLeuGlyLeuLeuLysAla

      3970                3990                4010
   TTTAACAACTTTCCAATCACTAATAAAATTCAATGCAACGGGTTATTCACTCCCAGTAAC
   PheAsnAsnPheProIleThrAsnLysIleGlnCysAsnGlyLeuPheThrProSerAsn

      4030                4050                4070
   ATTGAAACTTTATTAGGAGGAACTGAAATAGGAAAATTCACAGTCACACCCAAAAGCTCT
   IleGluThrLeuLeuGlyGlyThrGluIleGlyLysPheThrValThrProLysSerSer

      4090                4110                4130
   GGGAGCATGTTCTTAGTCTCAGCAGATATTATTGCATCAAGAATGGAAGGCGGCGTTGTT
   GlySerMetPheLeuValSerAlaAspIleIleAlaSerArgMetGluGlyGlyValVal

      4150                4170                4190
   CTAGCTTTGGTACGAGAAGGTGATTCTAAGCCCTGCGCGATTAGTTATGGATACTCATCA
   LeuAlaLeuValArgGluGlyAspSerLysProCysAlaIleSerTyrGlyTyrSerSer

      4210                4230                4250
   GGCATTCCTAATTTATGTAGTCTAAGAACCAGTATTACTAATACAGGATTGACTCCGACA
   GlyIleProAsnLeuCysSerLeuArgThrSerIleThrAsnThrGlyLeuThrProThr

      4270                4290                4310
   ACGTATTCATTACGTGTAGGCGGTTTAGAAAGCGGTGTGGTATGGGTTAATGCCCTTTCT
   ThrTyrSerLeuArgValGlyGlyLeuGluSerGlyValValTrpValAsnAlaLeuSer

      4330                4350                4370
   AATGGCAATGATATTTTAGGAATAACAAATACTTCTAATGTATCTTTTTTAGAGGTAATA
   AsnGlyAsnAspIleLeuGlyIleThrAsnThrSerAsnValSerPheLeuGluValIle

      4390                4410                4430
   CCTCAAACAAACGCTTAAACAATTTTTATTGGATTTTTCTTATAGGTTTTATATTTAGAG
   ProGlnThrAsnAlaEnd
```

FIG. 3

-106
MetSerLysThrThrLysLysPheAsnSerLeuCysIleAspLeuProArgAspLeuSer

LeuGluIleTyrGlnSerIleAlaSerValAlaThrGlySerGlyAspProHisSerAsp

AspPheThrAlaIleAlaTyrLeuArgAspGluLeuLeuThrLysHisProThrLeuGly

SerGlyAsnAspGluAlaThrArgArgThrLeuAlaIleAlaLysLeuArgGluAlaAsn

GlyAspArgGlyGlnIleAsnArgGluGlyPheLeuHisAspLysSerLeuSerTrpAsp

IleArgAlaThrGlySer+1MetGlyAsnSerGlyPheTyrLeuTyrAsnThrGluAsnCys

ValPheAlaAspAsnIleLysValGlyGlnMetThrGluProLeuLysAspGlnGlnIle

IleLeuGlyThrThrSerThrProValAlaAlaLysMetThrAlaSerAspGlyIleSer

LeuThrValSerAsnAsnSerSerThrAsnAlaSerIleThrIleGlyLeuAspAlaGlu

LysAlaTyrGlnLeuIleLeuGluLysLeuGlyAspGlnIleLeuAspGlyIleAlaAsp

ThrIleValAspSerThrValGlnAspIleLeuAspLysIleLysThrAspProSerLeu

GlyLeuLeuLysAlaPheAsnAsnPheProIleThrAsnLysIleGlnCysAsnGlyLeu

PheThrProSerAsnIleGluThrLeuLeuGlyGlyThrGluIleGlyLysPheThrVal

ThrProLysSerSerGlySerMetPheLeuValSerAlaAspIleIleAlaSerArgMet

GluGlyGlyValValLeuAlaLeuValArgGluGlyAspSerLysProCysAlaIleSer

TyrGlyTyrSerSerGlyIleProAsnLeuCysSerLeuArgThrSerIleThrAsnThr

GlyLeuThrProThrThrTyrSerLeuArgValGlyGlyLeuGluSerGlyValValTrp

ValAsnAlaLeuSerAsnGlyAsnAspIleLeuGlyIleThrAsnThrSerAsnValSer

PheLeuGluValIleProGlnThrAsnAlaEnd

European Patent Office

EUROPEAN SEARCH REPORT

Application Number

EP 91 10 6110

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int. Cl.5)
X	M. COMANDUCCI ET AL. 'Identification and characterization of a 28-kd protein of C. trachomatis encoded by the 7.5-kb common plasmid' 1990 , IN 'CHLAMYDIA INFECTIONS'. ED. W.R. BOWIE ET AL. , CAMBRIDGE UNIVERSITY PRESS, NY, US pages 121 - 124 * Whole article *	1-3,6, 14,20,24	C12N15/30 A61K39/118 G01N33/569 C07K15/00
X	EP-A-0 336 412 (SCLAVO S.P.A.) * Whole document *	1-19	
Y		20-33	
Y	INFECTION AND IMMUNITY vol. 57, no. 9, September 1989, US pages 2683 - 2690; S. BIRKELUND ET AL: 'Characterization of native and recombinant 75-kilodalton immunogens from Chlamydia-trachomatis serovar L2' * Whole article *	20-33	
X	PLASMID vol. 23, 1990, NEW YORK, US pages 149 - 154; M. COMANDUCCI ET AL.: 'Diversity of the Chlamydia trachomatis common plasmid in biovars with different pathogenicity' * Whole article *	1-19	TECHNICAL FIELDS SEARCHED (Int. Cl.5) C12N A61K C07K
X	MOLECULAR MICROBIOLOGY vol. 2, no. 4, 1988, OXFORD, GB pages 531 - 538; M. COMANDUCCI ET AL.: 'The structure of a plasmid of Chlamydia trachomatis believed to be required for growth within mammalian cells' * Whole article *	1-19	
Y		20-33	

-/—

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
BERLIN	19 MAY 1992	JULIA P.

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P0401)

European Patent Office

EUROPEAN SEARCH REPORT

Application Number

EP 91 10 6110

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int. Cl.5)
Y	INFECTION AND IMMUNITY vol. 59, no. 1, January 1991, US pages 79 - 90; M.C. CERRONE ET AL.: 'Cloning and sequence of the gene for heat shock protein 60 from Chlamydia trachomatis and immunological reactivity of the protein' * Whole article *	20-33	
X	NUCLEIC ACIDS RESEARCH. vol. 16, no. 9, 1988, LONDON, GB pages 4053 - 4067; C. HATT ET AL.: 'Analysis of the entire nucleotide sequence of the cryptic plasmid of Chlamydia trachomatis serovar L1. Evidence for involvement in DNA replication' * Whole article *	1-19	
Y		20-33	
Y	EP-A-0 293 079 (368800 ALBERTA LTD) * Whole document *	20-33	
			TECHNICAL FIELDS SEARCHED (Int. Cl.5)
X	NUCLEIC ACIDS RESEARCH. vol. 15, no. 24, 1987, LONDON, GB page 10596; K.S. SRIPRAKASH ET AL.: 'A gene for dnaB like protein in chlamydial plasmid' * Whole article *	1-4,12	
X	PLASMID vol. 18, no. 3, 1987, NEW YORK, US pages 205 - 214; K.S. SRIPRAKASH ET AL.: 'Characterization and sequence of a plasmid from the trachoma biovar of Chlamydia trachomatis' * Whole article *	1-19	
Y		20-33	

-/---

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
BERLIN	19 MAY 1992	JULIA P.

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P0401)

European Patent Office

EUROPEAN SEARCH REPORT

Application Number

EP 91 10 6110

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int. Cl.5)
Y	EP-A-0 192 033 (CHIRON CORPORATION) * Whole document *	20-33	
			TECHNICAL FIELDS SEARCHED (Int. Cl.5)

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
BERLIN	19 MAY 1992	JULIA P.

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P0401)